MPHIL

P-Glycoprotein Genes in Haemonchus Contortus

Kenworthy, James

Award date:
2013

Awarding institution:
University of Bath

P-GLYCOPROTEIN GENES IN *HAEMONCHUS CONTORTUS*

JAMES D. W. KENWORTHY

A thesis submitted for the degree of Master of Philosophy

University of Bath

Department of Biology and Biochemistry

May 2013

.....

ACKNOWLEDGEMENTS

Thanks to my supervisors

Dr Priest, Dr Wolstenholme, Dr Walsh and Dr Coles.

Thanks for all the support from my family and friends.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF SEQUENCES

TABLE OF EQUATIONS

LIST OF ABBREVIATIONS

ABC	ATP binding cassette
ASS	*assembled contigs* database released in November 2007
ATP	adenosine triphosphate
AV	avermectins
BAC	bacterial artificial chromosome
BAC	*BAC end sequence reads* database
BLAST	Basic Local Alignment Search Tool
BLASTP	BLAST for amino acid queries of amino acid subjects
BLASTX	BLAST for nucleotide queries of amino acid subjects
C09	*assembled contigs* database released in August 2009
CD	Conserved Domain
cd	conserved domain
CDD	Conserved Domains Database
COG	clusters of orthologous groups
CON	*assembled contigs* database released in January 2006
contigs	contiguous sequences derived from shotgun sequencing
C-terminus	carboxyl-terminus
EST	expressed sequence tags
EST	*expressed sequence tags* database
E-value	Expected value
FECRT	Faecal Egg Count Reduction Test
FOS	*fosmid end sequence reads* database
gDNA	genomic DNA
HSPS	high-scoring segment pairs
IBD	inter-nucleotide binding domains
IVM	ivermectin
kb	kilobase pairs
ML	milbemycins
ML	macrocyclic lactones
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
nr	non-redundant protein sequences
N-terminus	amino-terminus
pgp	P-gp gene
PGP	P-gp protein
P-gp	P-glycoprotein
PHYLIP	Phylogeny Inference Package
PSSM	Position-Specific Scoring Matrix
qPCR	quantitative real-time polymerase chain reaction
RPS-BLAST	Reverse Position-Specific Iterative BLAST for queries of PSSM
S	alignment score
S09	*supercontigs* database released in August 2009
Sanger	Wellcome Trust Sanger Institute
SEQ1	*sequence reads* database released in December 2004
SEQ2	*sequence reads* database released in August 2005
SEQ3	*sequence reads* database released in November 2005
SNP	single nucleotide polymorphisms
TBLASTN	BLAST for amino acid queries of nucleotide subjects
TMD	transmembrane domain
UNA	*unassembled reads* database released in January 2006
WU-BLAST	Washington University Basic Local Alignment Search Tool

SUMMARY

Control of *Haemonchus contortus* currently relies on anthelmintics including ivermectin (IVM). The existence of IVM resistance threatens the sheep industry. Detailed knowledge of the underlying genetic mechanisms may reveal opportunities to avoid selection of IVM resistance. Molecular tests are required in order to more sensitively identify and quantify resistance and then test competing hypotheses on the prevention and control of resistance. It is first necessary to identify those genes or loci under selection following IVM application and detect genotypic markers of IVM resistance.

However, as the genetic basis of IVM resistance is not yet established and in the absence of genomic tools currently available for *H. contortus*, the approach taken to date is the analysis of the gene families implicated in IVM resistance including the P-glycoproteins (P-gps). *H. contortus* has an incomplete genome without annotation and only one P-gp messenger RNA (mRNA) transcript has been fully sequenced. The aim of this project was to discover the incomplete and dispersed P-gp genetic sequences in the sequence databases and construct the putative genes.

Bioinformatics was used to reveal *H. contortus* sequences that map both from and to known P-gp sequences in *Caenorhabditis elegans* by a process of Basic Local Alignment Search Tool (BLAST) algorithms, protein distance matrices, phylogenetics and gene structure prediction algorithms. Evidence of up to 11 P-gp genes (*pgp*) was discovered, which is fewer than the 15 in *C. elegans*. There are eight clades in *C. elegans* (*Cel-pgp-1*; *2*; *3*/*4*; *5*/*6*/*7*/*8*; *9*; *10*; *11*; *12*/*13*/*14*/*15*) of which seven have representatives in *H. contortus*; there are no homologues to the entire clade of four genes *Cel-pgp-5* to *8* inclusive. All the duplications in *C. elegans* are on the X chromosome but none of these groups has more than one homologue in *H. contortus*. Conversely autosomal *Cel-pgp-2* and *Cel-pgp-9* with evidence of multiple homologues in *H. contortus* are not duplicated in *C. elegans*.

1 INTRODUCTION

1.1 Helminths

Helminths comprise trematodes (flat worms), cestodes (tape worms) and nematodes (roundworms). The Nematoda phylum (Diesing, 1861) as roundworms use their longitudinal muscles and high internal pressure to provide the thrashing motion of their entire body for locomotion. Larval and adult motility assays utilise this action to measure anthelmintic inhibition in sensitive organisms and test for successful movement even in the presence of anthelmintic in resistant nematodes.

Reproduction is sexual in the majority of nematodes and in the two organisms in this study. The varied progeny increases the chance in each generation for some individuals to be resistant or less sensitive to noxious stimuli such as anthelmintic. Further generations of meiosis, recombination and crosses between male and females that are resistant or less sensitive enables a Nematoda population to become resistant.

The Nematoda are abundant in terms of species but importantly for this study also have large population sizes in an environment such as thousands in one ruminant host. This opportunity for multiple sexual partners further enhances variability across the population in a host or environment. Therefore if a novel or intermittent challenge such as anthelmintic application is less that 100% efficacious the surviving population will be large and concentrated enough to reproduce successfully. The nematodes are noted for their wide range of habitats, some species are very specific and limited whilst others are capable of a variety of environments. The parasitic nematode in this work has a global range and is known to infect a multitude of hosts.

Nematodes are protected by a secreted cuticle that is sufficient protection for invading the digestive tracts in the case of parasites. The pathogenic worm in question is safely able to pass through the preceding three stomachs to burrow into and then return into the lumen of the acidic fourth stomach of ruminant hosts. This cuticle layer defines the life cycle as four moults occur in the egg and larval stages before the adult stage in both the model organism and pathogen.[1]

One of the important sense organs of Nematoda are papillae, these are chemoreceptors[2] and provide the input for moving away from a noxious chemical such as anthelmintic. This ability is used in an adult motility assay through a barrier away from the anthelmintic under investigation. The amphid organs and their internal molecular biology are heavily implicated in resistance to two main classes of anthelmintics. A particular feature of the Nematoda nervous system is a ring of neurones encircling the pharynx. This is significant in this research as the drug of interest has an association to this anatomy.

Table 1: Taxonomy of *Haemonchus contortus.*

Taxa	Name and Reference
Kingdom	Animalia
Phylum	Nematoda (Diesing, 1861)
Class	Chromadorea (Inglis,1983) [3]
Subclass	Chromadoria (Pearse, 1942)
Order	Rhabditida (Chitwood, 1933) [3]
Infraorder	Rhabditomorpha (Oerley, 1880) (De Ley & Blaxter, 2002) [3]
Suborder	Rhabditina (Chitwood, 1933) [3]
Superfamily	Strongyloidea (Weinland, 1858) [3]
Family	Trichostrongylidae (Leiper, 1912) [3]
Subfamily	Haemonchinae (Skrjabin & Schulz (or Schul'ts or Schul'tz), 1952) [4]
Genus	*Haemonchus* (Cobb, 1898)
Species	***H. contortus*** (Rudolphi, 1803) (Cobb, 1898 (type for genus))

The next taxonomic level of the organism in this study as summarised in Table 1 is the Chromadorea (Inglis,1983) class, which usually have elaborate and spiral amphids. The significance is that the anthelmintic under consideration is associated with a physical effect on the amphids. The sophistication in Chromadorea is a factor to consider in the sensitivity and resistance of the organisms to the action of the anthelmintic at and through this organ.

The pharynx is more sophisticated in Chromadorea than other nematodes. The sensitivity of this organ to the action of the anthelmintic being investigated and the potential of a decreased sensitivity and resistance at this location is of interest to this study.

The Rhabditida (Chitwood, 1933) order contains both free-living and parasitic species. In some cases both modes occur in the same species at different life stages such as the free-living second larval stage (L2) feeding on bacteria in faeces and the parasitic adult stage (L5) feeding on the blood of the host in the

studied organism. The beginning of the third larval stage (L3) in the parasitic taxa is infective to the final host, this stage actively climbs grass leaves to be ingested in the pathogenic worm in this work. The Rhabditida contains both the free-living model organism and parasite in this research. The amphids appear pore-like with small openings, this precise anatomy is relevant with regards to changes in this structure associated with anthelmintic mode of action and resistance.

The Trichostrongylidae (Leiper, 1912) family are parasites of herbivores and primarily of ruminants, and as such contain most of the parasites economically important to agriculture. The parasitic species in this work infects sheep and goats alongside other incidental hosts. Trichostrongylidae adults invade the abomasum (fourth stomach) or intestine, the final larval stages and adults of this parasite only occupy the abomasum. The eggs pass through the gut and as Trichostrongylidae juveniles cannot penetrate skin they must remain on the pasture until ingested hence completing and defining the life cycle. The variable proportion of the total population in the host and therefore exposed to anthelmintic compared to the remainder in the environment is relevant to the pressure for selection of resistance to the anthelmintic.

The Strongylus genus was erected with *Strongylus contortus* (Rudolphi,1803) but then altered to *Haemonchus contortus* (Cobb, 1898) as the type for the new *Haemonchus* genus. The number and synonymy of members has been in flux in the literature with 30 papers raising an additional 30 species, four *species inquirendae*, two varieties and five subspecies. The majority of these have been synonymised or removed later. The most recent work concludes the *Haemonchus* genus comprises 12 species with no *species inquirendae*, varieties or subspecies as detailed in Table 2. However, the species *Haemonchus santomei* (Gutteres, 1949) has not been dismissed nor investigated since first raised.

Table 2: *Haemonchus* Species.

Species	Reference
H. contortus	(Rudulphi, 1803) (Cobb, 1898 (type for genus)
H. longistipes	(Railliet & Henry, 1909)
H. placei	(Place, 1893)
H. similis	(Travassos, 1914)
H. mitchelli	(Le Roux, 1929)
H. bedfordi	(Le Roux, 1929)
H. vegliai	(Le Roux, 1929)
H. lawrencei	(Sandground, 1933)
H. okapiae	(van den Berghe, 1937)
H. santomei	(Gutteres, 1949) not dismissed nor included since
H. krugeri	(Ortlepp, 1964)
H. dinniki	(Sach, Gibbons & Lweno, 1973)
H. horaki	(Lichtenfels, Pilitt, Gibbons & Boomker, 2001)
Total	**12/13**

Members of the *Haemonchus* genus are all blood-consuming abomasal (true stomach) parasites of a diverse range of 46 genera of hosts including sheep, goats, cattle, deer, antelopes, camels and giraffes. The type species of *H. contortus* infects the vast majority of these, a total of 40 host taxa.

One of the most globally prevalent and economically important parasitic species in domestic animals is *H. contortus*. In sheep this parasite is the most important as feeding on blood in the abomasum can lead to symptoms of anaemia, lethargy, weight loss and even death in acute infections (haemonchosis).

1.2 Anthelmintics

Treatment and control of parasitic worms (helminths) is dependent on de-worming drugs (anthelmintics) as there are no other practical choices,[5; 6] to the extent that anthelmintics are the most valuable veterinary products constituting 40% of all livestock health expenditure.[7] One of the three classic anthelmintic classes is the macrocyclic lactones (ML) group that comprises the avermectins (AV) and milbemycins (MI). AVs are 16-membered MLs isolated in the 1970s from soil-dwelling *Streptomyces*. The eight natural AVs demonstrate broad-spectrum activity against ectoparasites and nematodes with avermectin $B1_a$ being the most potent.[8] Commercial introduction was in 1981 with ivermectin (IVM) as a mix of two synthetic chemicals, the majority being 22,23-dihydroavermectin $B1_a$.[9] The market for IVM has now grown to become the most commonly used drug in

agriculture, with a global consumption of US$ 1 220 million per annum.[7; 9] It is used for both preventative treatment (prophylaxis) and clinical treatment of ectoparasite infestation and endoparasite infection.[7]

1.3 Anthelmintic Resistance

Drug resistance in veterinary helminths is a worldwide and serious problem in agriculture as chemotherapy is essential for agricultural income and animal welfare in the absence of viable alternatives.[5] In particular the widespread resistance to the older anthelmintics including narrow-spectrum drugs and all three broad-spectrum classes is serious in sheep[10; 11; 12] and has increased the reliance on IVM. The emergence of IVM resistance in *H. contortus* is therefore a serious threat to the control of the pathogen[11; 13; 14] to the extent of threatening the entire sheep industry in Australia.[13]

The purpose of chemotherapy is for a drug to be presented to a target to cause an effect. The target and effect can be defined at any level ranging from molecule to organism to population. If an effect does not occur it could be due to the drug, the presentation or the target, and these need to be examined before resistance is declared erroneously. Differences or problems in the formulation, chemical stability or expiration of the drug are the first confounding factors to rule out. Secondly, differences or problems in the presentation including dosage, method of administration and pharmacokinetics (which varies between different host species) need to be eliminated. As parasites are not usually identified before applying anthelmintics, species or life stages may be present and survive as they are not targets for the drug. This example illustrates the fundamental difference between *susceptibility* and *sensitivity*. *Susceptibility* is the first term to be considered as *sensitivity* only applies in one of the two *susceptibility* scenarios. If, with no previous exposure to the drug, the drug cannot induce an effect from the target, the target is *insusceptible* only, this must be considered to avoid incorrect and misleading conclusions of resistance. An *insusceptible* target is innately not vulnerable to the drug due to factors present before any use of the drug including protection by the host, the life stage, behaviour, target concentration and distribution, drug metabolism and excretion. A target that is vulnerable to a drug is defined as *susceptible*. A *susceptible* target is innately vulnerable to the drug in at least one host, life stage and season but might be *insusceptible* in others. A

susceptible target is also *sensitive* if presently affected by the drug. A *susceptible* target has *decreased sensitivity* if, on exposure to the drug, acquired characteristics mean the effect has lessened or the same effect requires a higher concentration within therapeutic limits. Once a *susceptible* target has acquired *decreased sensitivity* to the point where a drug is ineffective within therapeutic doses the target is finally *resistant*.

The *in vivo* test forms the phenotypic definition of resistance, such that the reduction in the number of nematode eggs counted in the faeces following anthelmintic treatment is less than 95% relative to the pre-treatment count (Faecal Egg Count Reduction Test; FECRT).[15] At the present time resistance can be identified and measured on the basis of changes in phenotype such as decreased inhibition of hatching or feeding, or an increase in survival following exposure of *H. contortus* to IVM *in vitro*. These phenotypic tests are neither sensitive nor reliable enough to detect early selection of resistance but are restricted to being indicators of clinical resistance.[16; 17; 18]

1.4 Genetics of Anthelmintic Resistance

Parasites carrying mutations that decrease drug efficacy in that individual will provide a survival advantage in the face of anthelmintic treatment.[19] This selection for differences in the genetic profile is the first step in drug *resistance*.[20] The surviving male and female adults are able to reproduce and the genetic profiles are inherited by their offspring.[19; 21] In the case of allelic variation causing *resistance*, whether the alleles are recessive or dominant will affect whether heterozygotes or homozygotes progeny will also survive.[19; 22; 23] Another control on the genotype frequencies is that in some cases homozygotes are lethal whereas the heterozygotes confer the survival advantage.[19] If anthelmintic is then absent the changes are retained in this heritable shift, so creating a ratchet effect on resistance with an increase, retention then further increase.[24; 25; 26] With selection pressure continuing to be applied, the frequency of the resistant genotypes increases until a significant proportion of the population are resistant to therapeutically relevant doses of drug.[21] This boundary in drug resistance is pertinent with parasites as the dosage has to be below toxic levels for the host. At this stage treatment has failed and the population of parasites in that location is classed as *resistant* to that anthelmintic.[21]

Sequence variation in parasitic nematodes is high with evidence originally from allozyme analysis and then confirmed by restriction enzyme analysis, single-strand chain polymorphism and now direct sequencing of both mitochondrial and nuclear DNA.[27; 28; 29; 30; 31; 32; 33; 34; 35] The diversity at the nucleotide level is the standard measure for reporting[36] and is at a rate of 0.011 in *C. elegans* but as frequent as 0.026 in parasitic nematodes.[37] Base diversity is directly related to mutation rate and population size. A mammalian host such as sheep can be infected with thousands of adult trichostrongylid adult nematodes and one female can pass thousands of eggs per day.[38] Whilst the effective population size from a genetics standpoint is less than the total number of eggs produced, it is still extremely large in trichostrongylids compared to most other dioecious organisms.[21] Within the parasitic nematodes the group with the greatest variability are the trichostrongylids, including *Haemonchus*.[39]

Recombination is a further source of genetic variation correlated with physical distance between loci on the same chromosome, recombination rate per chromosome per meiotic event and the number of meiotic events, which is the number of generations and hence time divided by average life cycle period. Recombination rates are not known for parasitic nematodes.[21] Linkage disequilibrium is rapidly broken down by recombination in a large effective population and so linkage disequilibrium is inversely proportional to population size. Therefore in the abundant and dense population of trichostrongylid nematodes such as *H. contortus*, linkage disequilibrium is expected to be short lived.[21]

Recombination is also a cause of breaking down associations of genetic markers and alleles and thus generating linkage equilibrium. Where particular combinations of alleles are deleterious or beneficial their proportion in the population is less or more than expected by random recombination from the frequency of each allele in the population, creating linkage disequilibrium. In a population under anthelmintic treatment linkage disequilibrium is used as a signal of a genotype conferring resistance. Where a signal does not vary over a physical distance on a chromosome, linkage is present. This is the case in *H. placei*, which has almost complete linkage within 1 kb of a GluCl gene.[40] Those genetic markers in linkage or linkage disequilibrium in a population exposed to anthelmintic are evidence of selection at or near those loci.[41; 42]

A hard sweep describes a single mutation immediately before or following drug application causing resistance, which is strongly selected for and both the mutation and its associated genetic markers in the original haplotype rapidly rise in frequency in the population. Insufficient time has elapsed for recombination to break up the original haplotype and so linkage disequilibrium and a profound reduction in polymorphism extending out for a large segment from the resistance locus is detected.[21; 42]

A hard sweep has occurred in *Drosophila melanogaster*,[43] *Plasmodium falciparum*,[44; 45] rats[46] and in both the trichostrongylids *Teladorsagia circumcincta* and *H. contortus* a predominant haplotype for BZ resistance has been identified.[47]

A soft sweep is the selection of multiple different haplotypes conferring drug resistance and seems more likely than a hard sweep in anthelmintic resistance.[21; 42; 48; 49] Multiple haplotypes comprise a complex scenario where the genetic footprint is weak and hard to detect.[21] However, whilst the reduction in polymorphism is minimal, an unusually strong pattern of linkage disequilibrium is discernable.[49]

The origin or multiple origins of drug resistance could have been in existence prior to the use of anthelmintics, a novel change just prior to or during treatment, or recurrent mutations. In addition migration can introduce any of these three origins onto a farm or geographic area through movement of the livestock hosts. The origin of anthelmintic resistance is important as it affects which methods are most efficient in investigation, surveillance, control and treatment. The candidate gene approach would best suit a single genetic origin whereas genome-wide studies would assist with multiple or recurrent causes. Novel or recurrent origins would dictate that on-farm control is more important than quarantine measures, whereas pre-existing alleles that can be introduced by livestock are best protected against by biosecurity and treatment followed by testing for any remaining nematodes during quarantine.[21]

The case for a recent spontaneous origin of anthelmintic resistance is the plentiful and continual supply of mutations in nematodes with a large population and high mutation rate.[21] The mutation rate in nematodes is higher than other

phyla[30; 39; 50] in both mitochondrial and nuclear DNA in *C. elegans*.[51; 52] The polymorphism rate per base pair has been measured as 9 x 10^{-9} in *C. elegans* and calculated as 2 base pair substitutions per nuclear genome per generation[53] or one SNP for every base on average in 5 x 10^{-7} individuals. A sheep with a moderately high *H. contortus* burden would shed that number of eggs in a few days demonstrating the variability in the population. Data from the *H. contortus* genome sequencing project reveals insertions and deletions (indels) from a few bases to many kb are commonplace. The mutation rates for the model organism are underestimated as half of mutations are indels in addition to SNPs, therefore the potential for both single and multiple independent resistant alleles is evident.[21]

A direct demonstration of multiple independent origins of resistance in *T. circumcincta* and *H. contortus* is the presence of two different resistant alleles on one goat farm and alleles unique to different goat farms closed to livestock movement.[47] The case for pre-existing alleles lost by genetic drift in other farms is unlikely due to the parasite population size and BZ selection pressure on those other locations.[21] Current mathematical modelling also supports multiple independent origins especially with the large population size and high mutation rate in trichostrongylid nematodes.[48] More evidence for independent origins of BZ resistance is the difference in *β*-tubulin isotype 1 restriction fragment patterns between geographically separate *H. contortus*.[22]

The case for pre-existing alleles is strengthened by the ease and speed of selection for IVM-resistant *H. contortus* in just three generations.[54] The rapid emergence of tyrosine at codon 200 in *β*-tubulin isotype 1 preventing BZ binding and conferring resistance in *H. contortus* is most likely explained by pre-existing alleles at relatively high frequencies.[21; 23] The separate resistant alleles on the goat farms that suggest recurrent emergence are also consistent with pre-existing genotypes.[21; 47] The other argument from that work was the farms being closed to hosts meant the mutations had occurred independently is also explained by the sequence being present prior to closure.[21; 47; 55] The populations of *T. circumcincta* had little genetic differentiation between the farms, thus supporting the hypothesis that the same genetic origin was being registered.[21; 34; 56]

Migration of resistance has been proved with a single allele being discovered worldwide in insects.[57] With the tremendous amount of movement of livestock around the UK, the seven different resistant haplotypes in *T. circumcincta* in open UK flocks are postulated to be due to migration via ruminants.[21]

The number of stages and processes outlined belie the evidence that anthelmintic use inevitably and rapidly leads to resistance.[54; 58; 59] Anthelmintic resistance is most advanced in *H. contortus* and the understanding of the phenomenon in parasites is greatest for this species.[58; 59]

Table 3: Evidence of the Genetics of IVM Resistance.

Drug	Genetic Evidence	Species	Reference
IVM	ligand for P-gp	in mice	[60; 61]
IVM	inhibit transport function of some P-gp	in mice	[60; 61]
IVM & MOX	associated through population genetic studies with P-gp	in *C. elegans* & ***H. contortus***	[62; 31]
IVM	not consistent association through population genetic studies with P-gp	in ***H. contortus***	[63]
IVM	selection on ABC transporter	in *O. volvulus*	[64; 65; 66; 67]
IVM	decreased polymorphism of P-gp	in ***H. contortus***	[31; 68]
IVM	decreased polymorphism of P-gp	in *O. volvulus*	[69]
IVM	overexpression of P-gp	in ***H. contortus***	[70]

The genetic basis of IVM resistance in *H. contortus* is not currently known,[6; 71] which precludes any monitoring of resistant genotypes prior to phenotypic changes. The partial evidence of the genetics of IVM resistance is summarised in Table 3. One candidate are the P-glycoproteins (P-gp), these are members of the adenosine triphosphate (ATP) binding cassette (ABC) transporter superfamily and as such are efflux pumps in cell membranes, removing hydrophobic xenobiotics from the cell membrane and cytoplasm.[72; 73] There has been overwhelming evidence of their rôle in anthelmintic activity, IVM is a potent ligand and inhibits the transport function of some P-gps.[60; 61] P-gps have been associated with ML resistance from population genetic studies[31; 62] although this was not consistent.[63] However, in the nematode parasite *Oncophora volvulus* selection of IVM on an ABC transporter was verified[64; 65; 66; 67] and in both *O. volvulus* and

H. contortus IVM treatment led to decreased polymorphism of P-gps, a classic indication of selection.[31; 68; 69]

1.5 Study of Genetics of Anthelmintic Resistance

There are a number of approaches to studying the genetics and genomics of anthelmintic resistance as summarised in Table 4 and detailed below.

Table 4: Approaches to Studying Genetics and Genomics.

Option	**Alternatives**	
Relationship	Inheritance Linkage	Association
Region	Candidate	Genome wide
Variable	Allele	Expression
Discrete or Continuous	Case-Control	Quantitative
Outcome	Final	Intermediate
Trait	Phenotype	Genotype

An inheritance study investigates the genetic linkage to phenotype (not physical linkage on the chromosome between genes) through the comparison of parental genetics and their traits to that of the offspring. Classically, the link between candidate genes and outcome was established in this manner. However, genome-wide comparisons between parents and progeny have been made and could be for *C. elegans* in the lab and with sheep artificially infected with one male and one female *H. contortus*. This prevents use for investigating resistance in field populations under normal farming practice. Populations of known lineage with sensitive parental populations and resistant progeny in as little as three generations separation have been specifically retained as sources for inheritance linkage studies. The genetic linkage studies succeeded with single gene traits but were less reliable for multiple and complex causes of a phenotype.

Due to the limitations of genetic linkage studies, association between genotype and trait was initiated. As signals were expected from multiple genes combining in a complex interaction to produce an outcome the study of association was required. The populations are differentiated on trait with no knowledge of relationship or inheritance. The genotypes and traits of each individual are compared and the number of instances of each of the combinations of each genotype and traits established. The odds ratio of a genotype being present with a trait is calculated and the significant pairings are reported as associated.

Unfortunately, the vast majority of associations have low odds ratios, such that the genotype has a low effect on the trait being measured. This is especially the case when common polymorphisms are being investigated. Even in traits with known heritable components the association genes contribute relatively little. With a potentially limited contribution to the trait such as anthelmintic resistance, a limited contribution of genotype reduces the utility of such a finding as a molecular test of IVM resistance for example. As such the sensitivity is expected to be higher with fewer false negatives. In contrast the number of associations is more frequent and if association is carried out on a whole genome the number of false positives is increased and the specificity reduced. A limitation of association is that causality cannot be inferred and hence further more traditional and focussed molecular genetic work is required. However, it has a use as a first pass, particularly across the whole genome. Common polymorphisms are usually the subject matter in association studies and whilst these common alleles may be contributing, a more rare allele could not be included yet have an important contribution. Due to the restrictions of association, particularly of common SNPs, large numbers of organisms are required to provide sufficient sample sizes. With the population density of *H. contortus* this is not difficult but good definitions of cases and controls or of a quantitative trait are required.

Both association and inheritance linkage genotypes can be in the gene influencing the trait but it could also be close on the chromosome to the gene with a *rôle*, these two would be in physical linkage. Alternatively, the genotype could be in a non-random linkage with a locus not physically adjacent, if despite meiotic recombination the pattern remains and the two would thus be in linkage disequilibrium.

Populations should be controlled for sex, life stage and for population stratification from differing geographic sources. The genetic structure of the *H. contortus* global population is reported as a complex of biotypes, relevant to the population stratification that can be a confounding factor in association research.

A candidate gene approach uses existing knowledge to choose genes to specifically investigate as opposed to a whole genome approach. The knowledge can be sourced from a priori knowledge of the direct impact of a gene or gene family on a trait such as *β*-tubulin isotype 1 mutation with BZ binding and

resistance. This avenue can be limited so a more indirect approach is utilising known biological, physiological, functional or pathophysiological *rôles* of the genes that are relevant to the trait in question such as the IVM being a good substrate for the P-gp pumps. The sources may well be inconclusive but still point to genes of interest.

The obvious restriction to a gene being a candidate is reliance on previous knowledge and understanding of all the genetic contributions to a trait. Where multiple or complex interactions cause the trait an individual gene may not provide a clear association of inherited linkage. This can be counteracted by assessing several genes or families of genes simultaneously. Candidate gene studies often investigate sequence variation but can and do also measure expression.

Genome-wide or whole-genome studies consist of three main forms, studies of common variation, complete genome sequencing and microarrays. The majority of common variation is SNPs but such analysis requires knowledge of a substantial number of alleles and loci across the entire genome. By definition, sites of common variation will generate a large dataset across a genome and this sensitivity can provide opportunity for false positives. This method can be applied to association or inheritance linkage processes. In the case of association the number of results is impressive but the odds ratios can be less so.

One use for genome-wide sites of polymorphism is as a discovery cohort with no bias from pre-existing knowledge or assumptions. The alleles with the greatest and most significant odds ratios are then investigated further in a validation cohort. The use of a non-hypothesis-driven method is attractive in the case of IVM resistance as the progress in the gene families thus far has been limited.

With technological developments the alternative to partial sequencing is whole genome sequencing in high-throughput systems. The analysis of such data sets is an even more critical step. The first full genome is yet to be completed for *H. contortus* but high-throughput approaches can be applied to unknown genomes and the sequences then analysed on a project by project basis.

Microarrays detect changes in expression of transcripts genome-wide. The sensitivity to differences, number of results and the resulting potential for false positives are areas of concern.

Instead of a discrete variable being compared such as IVM resistance or sensitivity, a quantitative trait can be measured such as the concentration of IVM to cause an outcome or the measured motility in a set IVM concentration. The benefit of such an approach would be detection of intermediate populations with a decreased sensitivity but not resistant within therapeutic doses. This is useful in the study of hypotheses to control or prevent resistance and in surveillance where sub-clinical problems being identified could inform farmers and vets of the need to act. Where loci are being mapped with a quantitative trait the process is termed quantitative trait locus mapping, QTL mapping.

The trait being considered is often a phenotype such as the number of eggs, motility or development from one life stage to another. However, the outcome to compare can be a genotype. In this case variation in sequence or expression can be investigated with the expression of a particular transcript as the outcome. The process of mapping loci with the quantitative trait of the expression of a particular gene is termed expression quantitative trait locus mapping, eQTL mapping.

There are opportunities for non-candidate results to feed into candidate-driven work. Non-biased suggestions are then investigated more thoroughly. Genome-wide association studies can indicate associations between the trait and several possibilities. These new regions of interest can then be candidate loci and genes. The initial signal can be within a gene with potential influence or be in physical linkage with another gene. This physical linkage can be hypothesised when a genome provides whole chromosomal sequences, which is still to be completed for *H. contortus*. A linkage disequilibrium could also be the indirect connection and would need further investigation to decide. Quantitative trait locus mapping and microarrays can also distinguish sites as candidates for a causative role or being in physical linkage or linkage disequilibrium with a relevant gene or gene complex.

An expansion to the concept of the candidate gene being based on previous work is to be thorough in the bioinformatic analysis of the sequences metadata,

annotation and reported results across the genome, transcripts, genes and gene families to detect signals for candidate genes.

There are multiple areas requiring attention for the better prevention, detection and control of anthelmintic resistance. Tests need to be developed that are both reliable and practical in terms of equipment, time and cost in order to be used widely rather than when resistance is already established. The genetic mechanisms should be elucidated so that molecular tests can be developed that are sensitive to population changes prior to clinical manifestation. More generally an improvement is sought in the knowledge and understanding of the biology, ecology and epidemiology of parasites that could then allow for better interventions and provide conclusive answers to some of the debates surrounding parasite control. Finally, encouraging and creating systems and communication channels of expertise would allow these new lessons to be disseminated.[6]

1.6 Aim of Project

As described above, one of the areas requiring attention for the better prevention, detection and control of anthelmintic resistance is the elucidation of the genetic mechanisms so that molecular tests can be developed that are sensitive to population changes prior to clinical manifestation. In the absence of genomic tools currently available for determining genes under IVM selection in *H. contortus*, the approach taken to date is the analysis of the individual genes implicated in IVM resistance including the P-gp family. *H. contortus* has an incomplete genome without annotation and only one P-gp mRNA transcript has been sequenced.[70] The aim of this project was to discover the incomplete and dispersed P-gp genetic sequences in the sequence databases and construct the putative genes. These could then be investigated in the future for their rôle in IVM resistance in *H. contortus* and potentially utilised in molecular tests for IVM resistance.

2 MATERIALS AND METHODS

2.1 Bioinformatics

The purpose of bioinformatics was to discover novel genetic sequences in the incomplete *H. contortus* genome that map both from and to known P-gp sequences and therefore form putative *H. contortus* P-gp genes (*pgp*).

C. elegans and *H. contortus* are both in phylogenetic clade V of the nematodes[74] and *C. elegans* is commonly used as a model organism for *H. contortus*. The genome of *C. elegans* having been completed and annotated allowed extraction of the entire annotated *pgp* gene family from *C. elegans*, which was used to expand the P-gp protein (PGP) family in *H. contortus* from the single member (previously termed *Hco*-PGP-A) based on messenger RNA (mRNA).

Multiple genomic DNA (gDNA) databases existed for *H. contortus* as the nucleotide sequences were generated using six different methods and batches of new sequences released at seven different times between December 2004 and August 2009. There were a total of 21 separate searchable databases, comprising 12 distinct types of sequences with the remaining 9 databases having been included in and superseded by more recent collections.

2.1.1 Alignments along Entire Gene

The amino acid sequences for the PGP family in *C. elegans* of 14 proteins and the translation of the single pseudo-gene (*Cel-pgp-15ps*) were retrieved from the WormBase database (www.wormbase.org). These sequences were used to trawl the incomplete *H. contortus* genome databases hosted at the Wellcome Trust Sanger Institute (Sanger) for similar sequences using the Washington University-Basic Local Alignment Search Tool (WU-BLAST) (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/h_contortus). The specific WU-BLAST algorithm used was TBLASTN, in order to search translated nucleotide databases with the protein queries. (TBLASTN translated the nucleotide databases into all six reading frames and then used these amino acid sequences as the subject for a search using the amino acid sequence entered by the author.) The matrix used for amino acid substitution scores was the standard BLOSUM62.

The biological interest in an alignment was whether it provided evidence of homology between the underlying genes, however, the question has to be inverted, as is common in statistics, into the question of how much evidence existed that the alignment was due to chance alone. Continuing that logic meant that an alignment expected to occur less frequently by chance in comparison to another alignment was more likely to be homologous and biologically meaningful. To report that homology and significance, each alignment had an Expected value (E-value) calculated as below.

The query sequences are the amino acid identities of the known *C. elegans* and *H. contortus* proteins plus the translated pseudo-gene *Cel-pgp-15ps*. One amino acid sequence is compared to the subject database of *H. contortus* nucleotide sequences translated into all six reading frames. The amino acid residues are compared with a particular scoring matrix, BLOSUM62. Identical residues receive a highly positive integer score, residues with similar physico-chemical properties receive a moderately positive integer score and pairs of residues with differing properties are scored negatively. The BLOSUM62 scores were originally chosen to provide a negative sum for lengths of residues with no structural or functional similarity and thus a random alignment provides a negative total and is discarded.[75] A word length of 3 was used to seed alignments, for those that exceed a minimum positive score the alignment is tested whether extending in one or both directions increases the total score, this can involve a shorter length of fewer negative pairings in which case the alignment is trimmed to obtain the maximal score (S).

S

The resulting alignments that passes a minimum total score threshold are retained as the high scoring segment pairs (HSPs), the remainder are discarded. Each scoring matrix has a particular statistical effect on how the resulting scores compare to the minimum amount of information that is required to convey a range of alignments. This value for BLOSUM matrices is calculated in advance on random alignments and has a set value for alignments on a particular database. This number enables different scoring systems to be compared and is termed *lambda* (λ).[76]

$$\lambda \cdot S$$

The total score for an alignment multiplied by *lambda* is the exponent for calculating the number of alignments to be expected from chance alone.

$$\exp(\lambda \cdot S)$$

The exponent is negative such that a large positive score generates a small number of alignments expected from chance, a doubling in the score reduces the expected number. The expected value therefore approaches zero and frequently is reported as zero but is in fact always at least an infinitesimally small positive value.

$$\exp(-\lambda \cdot S)$$

The exponential function makes sense as for a score to double is the same as half the score occurring twice on the same alignment. The occurrence of that scenario due to chance alone is exponentially less likely than the best alignment only having available the score once.

The length of the subject database plays a role in the chance of an alignment as a larger set of sequences or a longer average sequence increases the chances in a linear function of an alignment with a certain score or above. A doubling of the total search space doubles the number of alignments with a certain score due to chance and hence doubles the expected number of alignments. For protein sequences the length of sequences in the database is limited to the range of protein lengths and is not arbitrary. The assumption in BLAST is that the lengths of the component individual sequences within the massed total length (n) is relevant.[77]

$$n \cdot \exp(-\lambda \cdot S)$$

$$n^{(-\lambda \cdot S)}$$

However, for genomic DNA databases the individual sequence lengths are arbitrary. The only practical calculation for calculating the number of expected

alignments to a DNA subject database is that the total massed sequence length is relevant and the individual sequence lengths are irrelevant. This holds true for very long gDNA lengths where the individual sequences lengths approach the maximal quantities. It is worth noting that for this research utilising on-going data releases of DNA databases the sequence lengths changed radically from thousands of bases in the *contigs* of 2006 (CON) to the hundreds of thousands of bases in the *supercontigs* of 2009 (S09). The calculated expected number of alignments was increasingly accurate.

The high scores are less likely near the ends of a sequence as the missing alignment reduces the positive score. This edge effect is minute in full genome coverage but increasingly pronounced with the more disjointed earlier databases.[76]

The characteristic of a database in producing expected numbers of alignments from chance is captured by the parameter K. Multiplying this with the expected number enables the appropriate scaling of the expected number and allows comparison between different databases.

$$K \cdot n^{(-\lambda \cdot S)}$$

The length of the query sequence (m) is clearly relevant to how many alignments would be expected from chance. Increasing the query length provides more lengths of sequence that could align with a particular score or higher. This relationship is such that doubling the length doubles the expected number of alignments.

$$K \cdot m \cdot n^{(-\lambda \cdot S)}$$

This function now generates the expect value (E-value) of the number of alignments with the particular score (S) or higher, taking into account the characteristics of the scoring system (λ) and subject database (K), the length of the query sequence (m) and the total massed length of the subject database (n).[78]

$$E\text{-value} = K \cdot m \cdot n^{(-\lambda \cdot S)}$$

Equation 1: E-value for Alignments.

Those seeded words exceeding the minimum initial score, then extended to obtain a maximal score locally, then exceeding the minimum score are the HSPS and those HSPS with an expect value exceeding the threshold are reported in the BLAST results.

The E-value (incorporating a capital E) was provided in shortened exponential notation (incorporating a lower case **e** to represent **e**xponent) e.g. 2e-3 equals 2 times 10 raised to the power of -3. This E-value was used to rank the alignments and thus the degree of homology between genes from other species and the *H. contortus* sequences: the closer the E-value was to zero the greater the homology inferred. The *H. contortus* sequences that had greater homology in the *back BLAST* with genes outside the *pgp* family were discounted in the search for the *H. contortus pgp* gene family, this removal of noise is the benefit of a *reciprocal BLAST*. Those sequences that matched translated *pgp* genes better than other gene families were putatively identified as components of the *H. contortus pgp* genes.

2.1.1.1 Alignments from Assembled *Contig*s and Expressed Sequence Tags Databases

In this first phase of bioinformatics the two sources most likely to reveal substantial information were investigated; these were the *assembled contigs* database (contiguous sequences derived from shotgun sequencing; released on 27 January 2006 and the most recent at the time; CON) and the expressed sequence tags (EST) database. The *contig*s database contained short fragments of 1-3 kilobase pairs (kb) length and the ESTs were typically only several hundred base pairs long. The results were returned with amino acid sequences from *H. contortus* (as translated from nucleotide sequences) aligned with amino acid sequences from *C. elegans* (as entered from proteins or the translated pseudo-gene). The alignments generated by WU-BLAST between *C. elegans* and *H. contortus* from the Sanger databases contained both relevant and irrelevant matches. To test whether a match was from a relevant putative gene, a second search in the opposite direction was utilised and termed a *back BLAST*, the combined searches

were described as a *reciprocal BLAST*. The *H. contortus* sequences were manually searched for highly similar or identical regions as the aim at this stage was to discover short identifying fragments for every gene rather than the entire gene sequences. These short amino acid sequences were provided as the query for *back BLAST* searches against the nucleotide collection database without restriction to any species; the program and database were hosted at the National Center for Biotechnology Information (NCBI; blast.ncbi.nlm.nih.gov/Blast.cgi). In order to ignore the expected differences in DNA for the same amino acid sequence between the species the *H. contortus* translated nucleotides were used as the query and a translated nucleotide database were used as the subject for these searches, which therefore required the TBLASTN algorithm. As the two databases investigated initially did not return homologous sequences long enough to be useful the other databases were investigated.

2.1.1.2 Alignments from Eight Databases

The next phase of bioinformatics compared the maximum length of alignments returned and coverage of the genome from all eight *H. contortus* databases available at Sanger at the time: the three *sequence reads* databases (released in December 2004, SEQ1; August 2005, SEQ2; November 2005, SEQ3); the *unassembled reads* database (released in January 2006; UNA); the *assembled contigs* database (released in January 2006; CON); the *BAC end sequence reads* database (bacterial artificial chromosome; BAC); the *fosmid end sequence reads* database (FOS) and the *expressed sequence tags* database (EST).

A larger batch of sequences was released as the *assembled contigs all reads* database (ASS) in November 2007. This database was queried with the entire *C. elegans pgp* gene family using TBLASTN. For the *back BLAST* the BLASTX program was run with the nucleotide sequences (translated by BLASTX) from *H. contortus* as the query against the non-redundant protein sequences (*nr*) database without restrictions to species of origin (program and database hosted at the NCBI) and thereby completed the *reciprocal BLAST* method. The E-value of aligned sequences was utilised to assess homology with *C. elegans* PGP proteins. More detailed and sensitive alignments were obtained using the BLASTX version of the BLAST2 program at NCBI, these were manually edited and cropped. An overlap between two sequences could have been insufficient for genome analysis

to categorise the pair as contiguous. However, the terminal sequences could have been sufficient for more general bioinformatic tools to detect an overlap. Hence both the Contig Assembly Program (CAP3), a component of the Mobyle platform (mobyle.pasteur.fr), and the BLAST2 program at NCBI was run on the entire set of *contigs* to detect overlaps, even if incorporating single nucleotide polymorphisms (SNP), as a precautionary measure.

2.1.2 Alignments to Specific Domains

Following the BLAST attempts at alignments, an alternative approach was required. Despite the presence of polymorphisms the P-gps, in common with all proteins, contained functional residues that were highly conserved. In addition, whole regions have particular functions and the range of variation at each residue is limited if the function of that domain is to be retained and for there not to be a fitness and survival penalty. This was exploited in the alignment of genomic sequences to known domains.

2.1.2.1 Alignments to Inter-nucleotide Binding Domains

Several sequences were known for the inter-nucleotide binding domains (IBD) in *H. contortus* and were used to solve the gDNA sequences in that region. Prior to use in *H. contortus* the process was evaluated for specificity with comparison of the results between amino acid and nucleotide investigations for *C. elegans*. The *C. elegans* amino acid sequences from each IBD were compared via the BLASTP software (designed for amino acid queries of amino acid subjects) at NCBI with the full-length *C. elegans* PGP protein sequences then the IBD amino acid sequences were aligned by TBLASTN to the translated nucleotide *pgp* sequences. The alignments and E-values were analysed for perfect matches and therefore confirmed the system could be trusted for aligning *H. contortus* sequences. The nucleotide sequences at Sanger were interrogated with the *H. contortus* IBD amino acid sequences by TBLASTN and both the text alignments and the E-value employed to determine which IBD amino acid sequences were represented within the *assembled contigs all reads* database (ASS). To corroborate the alignment the aligned sequences were assessed in their pairs via BLAST2.

2.1.2.2 Alignments to ATP Binding Cassettes

Further investigation of the possibilities of domain matching as an improvement on simple alignments included using the ScanProsite tools to probe the PROSITE database (provided by ExPASy; www.expasy.org/prosite). The next step was the exploration of the Conserved Domains Database (CDD) by means of the Conserved Domain Search (CD-Search) software at NCBI (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). A conserved domain (*cd*) record *cd03249* specific to P-gp was then utilised in the next phase. Whilst CD-Search was indispensable as a sensitive detection tool, the fact that *cd03249* was a Position-Specific Scoring Matrix (PSSM) and could be depicted using the PSSM Viewer at NCBI (www.ncbi.nlm.nih.gov/Class/Structure/pssm/pssm_viewer.cgi) revealed necessary detail of the individual scores and hence how fitting a residue was at each position. Furthermore, the PSSM Viewer permitted alignment of amino acid sequences to the PSSM and was able to generate E-values. This approach was the most apt for the P-gp family as synonymous SNPs were disregarded and non-synonymous SNPs were assessed according to the matrix of scores depending on the specific position of the variant amino acid. This analysis was considered to come closer to a test of function and so more reliable as an indicator of gene identity than the direct study of sequence.

The translated *contigs* previously extracted from the *H. contortus* ASS database and putatively aligned with *C. elegans* were aligned with *cd03249* using PSSM Viewer. These multiple alignments were arranged in parallel to both contrast the residues present at the same positions and also to gauge the number of domains present and hence half that number of *pgp* genes. The Jalview alignment editor was utilised for further manipulation of the *C. elegans* and *H. contortus* domains. The homology between the putative and annotated genes was approximated by a phylogeny inferred through computation of individual distance matrices (Jones-Taylor-Thornton model) for each *H. contortus* sequence with all the *C. elegans* ABC domains using the *protdist* component of the Phylogeny Inference Package (PHYLIP) hosted at Mobyle (mobyle.pasteur.fr/cgi-bin/portal.py?form=protdist). The initial alignments were replicated by 1 000 boot-strapping cycles and a distance matrix calculated for each in order to measure the robustness of the neighbour-joining consensus tree that was generated by the *neighbor* program offered at the same Mobyle site. This evidence of homology was used to group

the *contigs* according to *C. elegans* homologue. The disjointed sequences were linked into putative genes by inspection of the adjacent ends for evidence of a clean join but more commonly the intron-exon boundaries required correction and small gaps still persisted. At this juncture the *H. contortus* sequences from the ASS database composing the ABC had been detected and aligned.

2.1.2.3 Alignments to Both ATP Binding Cassettes and Transmembrane Domains

The method of using the *cd03249* PSSM to detect the ABC domains had revealed a portion of the proteins but pairing all the domains had not been possible. The method was therefore repeated with CD-Search employed to reveal other conserved domains specific to P-gp that were then aligned by the PSSM Viewer. A conserved domain (*pfam00664*) for the ABC transporter transmembrane domain (TMD) and for the entire multidrug resistance protein (a P-gp) in *Plasmodium falciparum* (*PTZ00625*) were investigated but a more sensitive matrix was discovered in the ABC-type multidrug transport system model (*COG1132*) that represented both the ABC and the TMD. This combined matrix meant those *H. contortus* sequences from the ASS database already aligned with the ABC domain were able to be extended into the TMD and additional sequences were detected that only aligned within the TMD. However, full resolution of the TMD was restricted due to the lower conservation in comparison to the ABC.

This method was repeated with the longer *contigs* from the *assembled contigs* database (C09) when it became available in August 2009. With longer sequences the benefit was greater than just the proportional increase in *contig* length or percentage coverage of the genome. This was because a greater proportion of the longer sequences covered both the ABC and TMD, even if only partially, and therefore the PSSM Viewer was more sensitive and able to increase the coverage of the *pgp* gene family. However, the C09 database was still too fragmented to provide full-length alignments and gaps remained.

The *supercontigs* sequences in the S09 collection comprised *contigs* of known length and nucleotide identities (present in the C09 database), separated by regions of known length but unknown sequence, designated by strings of *n*. This evidence removed the previous necessity of individually determining which *contigs*

were adjacent by manual methods as with the ASS *contigs*. Due to the length of these sequences (despite the unidentified regions) alignment to full-length genes was again attempted, similar to the original method with the much shorter sequences from the ASS database.

2.1.3 Alignments along Entire Gene

The 16 known PGP amino acid sequences (*Hco*-PGP-2.1.A, *Cel*-PGP-1 through 14 and *Cel-pgp-15ps*) were queried against the *supercontigs* database (S09; released in August 2009) using WU-TBLASTN, hosted at Sanger. The sequences with E-values below 10 were extracted from the database; although this E-value was high and allows false positives (lower specificity) it was also less likely to cause false negatives and so was more sensitive. Every putative *pgp* sequence from this *forward BLAST* was then compared to the *C. elegans nr* database. Sequences with at least regional matches (higher S scores; lower E-value) to *pgp* rather than other gene families were considered to contain *pgp* exons. The sequence was then compared to the amino acid sequences (*Hco*-PGP-2.1.A, *Cel*-PGP-1 through 14 and the translation of *Cel-pgp-15ps*) in order to determine putative homologues. Whilst the reciprocal BLAST methods were more sensitive to these sequences of greater coverage, the need for manual editing of the alignments and determining misalignments, particularly around the intron-exon boundaries and the unidentified regions was still necessary.

2.1.3.1 Gene Structure Prediction Algorithms

The alternatives to BLAST and manual editing included mGene, Splign, Spidey, DiAlign and GeneWise; all were evaluated for use in predicting the *pgp* gene sequences in *H. contortus* but only the GeneWise program was productive in this aim.

The purpose of GeneWise was as a comparison of genomic sequence with a homologous protein in order to predict the gene sequence and structure. The GeneWise method integrated two different processes sequentially, the first being a model of alignment of a homologous protein to the translated genomic sequence and the second is a model of predicting the gene structure.

All possible translations of the genomic sequence in all reading frames were considered against the homologous protein in order to generate a predicted amino acid sequence. Accepting all reading frames at every base allowed for insertion or deletion errors to not cause the remainder to be out of phase. These nucleotides were taken account of by a codon being allowed to be 2, 4 or 5 bases where the codons upstream and downstream provide a better alignment. However, no attempt was made to translate the abnormal codon because the calculation of all possibilities is too complicated and computationally expensive. The codon thus provided a null residue. This approach was far more resilient to resolving a genomic sequence to exons that pass over errors.

The alignment model does not attempt to locate substitution errors originating from the genomic sequencing. Instead the probability of a transition or transversion was included within the calculation for a triplet to be aligned with an amino acid. For those triplets where a reported base was different from the organism the effect on the probability assigned to a particular translated sequence and the scoring of that alignment was spread over a wider locus than the actual nucleotide. This smudged effect is of no consequence to the predicted gene structure.

GeneWise did not itself provide the profiles of protein homology for the model to use but instead incorporated an established *'worm'* pattern. The protein model was affected by three considerations: amino acid bias in the protein provided, codon bias in the organism providing the genomic material and sequencing error in the determination of that genomic sequence. The amino acid bias and codon bias are incorporated into the probabilities being calculated for a translated triplet being aligned with a base. The general 'worm' model provided the bias parameters and the default sequencing error rates of 1×10^{-5} was used as it is the reported standard for genome sequencing.

The second part of the amalgamated algorithm predicted gene structure. The aligned protein, splice site patterns and parameters for the genomic sequence fed into a maximum likelihood estimate of the various splice sites along the DNA sequence. A related approach from the protein alignment model was applied in the gene prediction model for splice sites in phase 1 and 2 that therefore split a codon by the intron. The signal from both 5' and 3' splice sites were received and contributed to the predicted gene structure. The translation of the codon was then

ignored to avoid a further submodel for just these calculations of identity. The position in the alignment over the splice site codon was not scored to protect against a negative score when the amino acid cannot be aligned to a non-diagnosed identity. A negative score would have reduced the evidence for this alignment and the overall score, in the case of marginal alignments this could have swung the evidence to report alternative exons or end the alignment early.

A number of the protein alignment model and gene structure model outputs were not possible, for example nonsensical genetic sequence. These states were forced to zero and then removed. Also some theoretical intermediate or final cases were not required, for example concerns of multiple changes occurring in one codon where the greatest portion of the possibilities have infinitesimally small probabilities. These were also forced to output a probability of zero without calculation and hence the states are not present in the process.

The gene structure algorithm had the ability to distinguish further signals of the central intron, polypyrimidine tract and the spacer at the end of the intron. The sequential application of the homologous alignment and the gene structure models including the intron signals would have resulted in 21 states and 93 transitions in total. That number of calculations was ambitious and slow.[79] More significantly, the intron signals were too sensitive to general gene structure in the genomic sequence provided despite the homology being distant and the alignment poor. Determining the exact position of an indel sequencing error where a 2, 4 or 5 base codon has been utilised was also too expensive. The error locus was likely to be 1 position different from that reported by the model but that was of no consequence. Losing the intron and exact error positioning states and transitions brought the combined model down to a manageable 6 states and 23 transitions. This version is termed GeneWise 6:23 or 623 in the parameters reported in the results and appendix and was the version employed in this study. The functional use of this model has been reported as being excellent in previous studies with the edge detection of exons providing a strong signal principally through the splice sites.

There were several modes available in GeneWise to account for different alignment and gene structure requirements. In the expansive model of 21 states and 93 transitions all genes score well. Mispredictions could therefore occur both in terms of any alignment to non-homologous proteins but also extended coverage

of the genomic sequence with the predicted gene, especially with the terminal exons being predicted for the best start and stop signals and splice sites. A specific flanking model as an addition to the combined model was too expensive. The global mode did allow for stretched genes but was useful as an initial choice to detect the core alignment. The wing approach was to not penalise alignments that start or end within the terminal five bases and hence allowed truncation and downstream starts and upstream stops, this mode was not more effective than other modes and was not utilised. Endbias was employed for genomic sequences expected to contain the full transcript and was helpful due to the tails being less well conserved and hence this choice aided in establishing the regions of lower identity or similarity as still being homologous.

The specificity of GeneWise has been reported as 87-96%, dependent on phylogenetic distance. Patterns in the false positives were low complexity regions and alternative splicing.[79] Alternative splicing is possible in the *H. contortus* P-gp and could have generated spurious matches but the model was robust enough. In this study the sequences had already been through a selection process and so specificity was less of a concern.

The sensitivity of GeneWise was relatively low in previous studies with only 60% of known transcripts being detected if the input is a short protein or a fragmented amino acid sequence. With the long and full-length P-gp sequences from *C. elegans* this was unlikely to be a problem. In other cases the sensitivity was reported to rise to 98% when the similarity of the amino acid sequences was 85-95% through to within 20 bases of the termini.[79] The incomplete *H. contortus* genome and hence fragmented genes was a potentially limiting factor in the sensitivity.

All sequences from the S09 database that had previously been aligned with *pgp* were re-evaluated using this GeneWise algorithm. The genomic sequences were entered as the target and the homologous amino acid sequence entered as the query into the *GeneWise* component of the WISE2 package hosted at Mobyle (mo byle.pasteur.fr/cgi-bin/portal.py?form=wise2). The comparison matrix used was BLOSUM62 and the gene parameter file selected was *worm*. The type of match attempted was variously *global* or *local* dependent on which predicted the longest gene structure with the least artefacts. Likewise the figures used as gap penalties

and gap extension penalties were altered from the defaults of 12 and 2 to provide better results as necessary. The substitution and insertion/deletion error rates were rarely altered from the defaults of 10^{-5}. All parameters have been provided at the start of every result, included in the Appendix. For each putative *pgp* sequence the predicted alignment, cDNA and amino acid sequences were generated; both the positions and sequences of the introns and the exons were provided. The known lengths of unknown sequence, denoted as strings of *n*, in the *supercontigs* did cause artefacts in the GeneWise outputs as inappropriate cut-offs in exons or were translated into unknown residues, denoted as *X*. Longer gaps in some *supercontigs* prevented resolution until the sequence was analysed in two sections and then manually edited.

3 RESULTS

3.1 Alignments along Entire Gene

3.1.1 Alignments from Assembled *Contig*s and Expressed Sequence Tags Databases

The first stage of the bioinformatic approach taken in this research is of querying the incomplete *H. contortus* genome, via the multiple databases becoming available over the course of the research, with known *Cel*-PGP. This produces high-scoring segment pairs (HSPS) with an example in Figure 1.

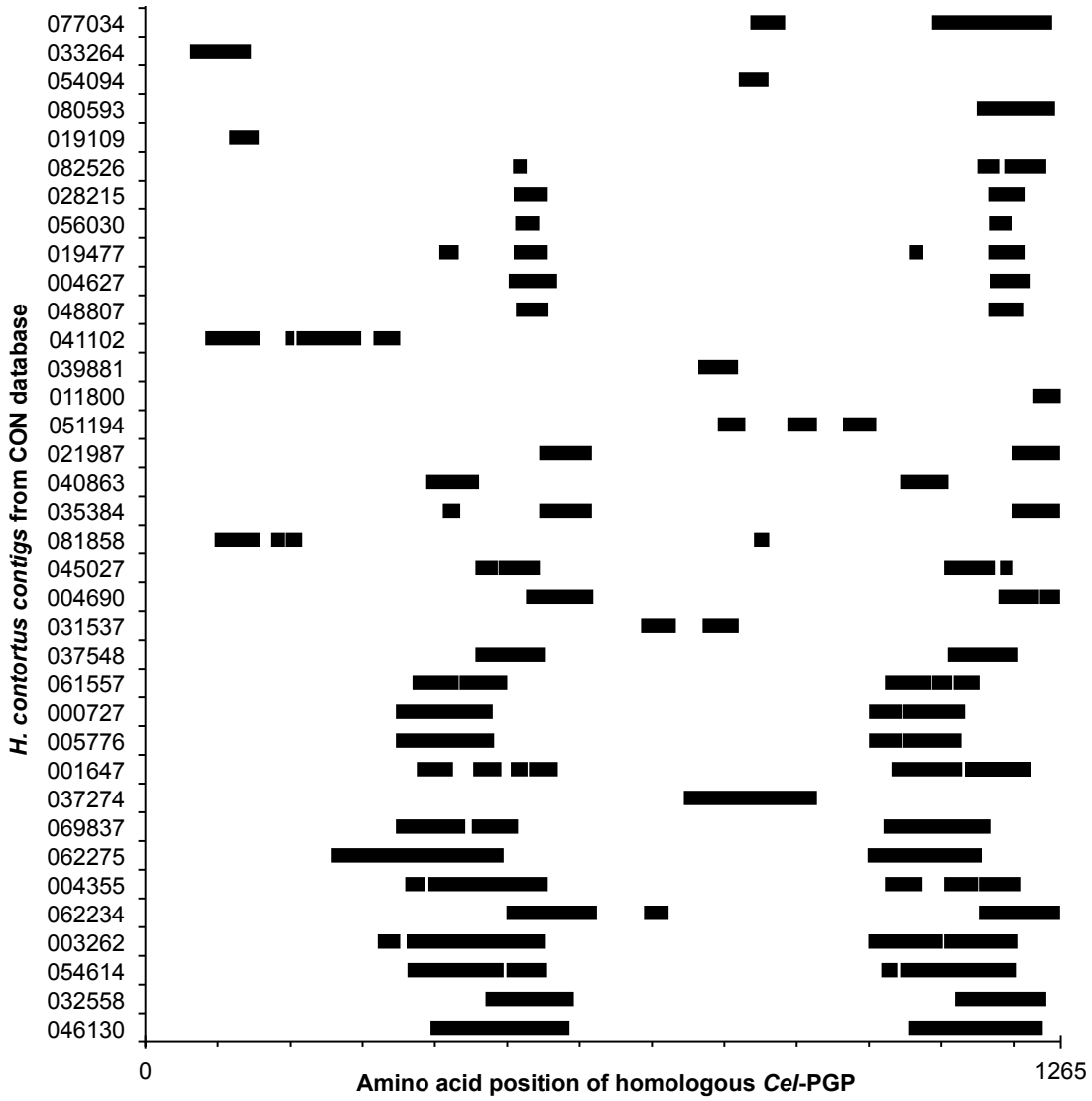


Figure 1: Alignments of *H. contortus* Sequences from CON Database with *Cel*-PGP-4.

Residue position of *Cel*-PGP-4 along *x* axis. Alignment of HSPs from both the forward and complementary sequences from CON database. Separation between matches indicates the fragmented short genome coverage as exons and separating introns are included in the bars. Marked increase of alignments at the ABC domains both 5' (around residue position 500) and 3' (around residue position 1150) indicate the conservation of these domains and alignments from genomic sequence homologous to other members of the *pgp* family.

The CON database of assembled *H. contortus contig*s released in 2006 revealed multiple matches on searching with BLAST. Figure 1 illustrates the HSPS matching *Cel*-PGP-4, and is representative of the other 14 members of the P-gp family in *C. elegans*. The *contig*s had multiple HSPS therefore the number of HSPS in Figure 1 is greater than the number of *contigs*, this is also demonstrated in Figure 2.

Examining the distribution of the HSPS in Figure 1 clearly shows the cluster around amino acid position 350-600 and from 1000 to the carboxyl-terminus (C-terminus), these are the two ABC and are highly conserved enabling highly scored alignments. This increases the sensitivity of this method to detect homologous sequences in these regions. This figure is similar to the graphic output from querying the other databases, although the number of HSPS varies. The later databases show increasing numbers of HSPS and greater coverage of the *pgp* family.

3.1.2 Alignments from Eight Databases

To visualise the extent of coverage of the *H. contortus* genome by the various Sanger databases the aligned *H. contortus* regions were depicted on a floating bar chart in Figure 2. The positions of these *H. contortus* sequences within each putative *H. contortus* PGP protein sequence were by analogy with the annotated locations of the homologous *C. elegans* sequences within the relevant *C. elegans* PGP protein. Figure 2 is an important chart as it illustrates the significant differences between the earliest databases in terms of the length and number of sequences returned, which is summarised in Table 5. The FOS database produces no *pgp* sequences. The BAC database gives only one relatively short sequence. The UNA source reveals matches along most of the length of *pgp* but with none overlapping and therefore provides no evidence of more than one *pgp*. The SEQ databases (SEQ1, SEQ2 and SEQ3) show increasing coverage with later releases as is expected from better and more thorough genome coverage. The CON database provides the greatest number of alignments at this stage, which were grouped in terms of which *Cel*-PGP they were most homologous to and these were then plotted in separate rows in the bar chart, one of these being the known mRNA and protein having been named *pgp-A* and PGP-A respectively. This early evidence points to ten *Hco-pgp* but later evidence in

Figure 8 is for nine *pgp* in *H. contortus*. However, the alignments these scores are calculated on are very short with just a few dozen amino acids from a couple of hundred nucleotides. The short white gaps in the bars indicate lack of alignment at the start and end of exons and are artefacts from the BLAST method rather than missing sequences. Following on from the CON database, the improved *assembled contigs* database (ASS; released in November 2007) containing longer

sequences is investigated in Figures 3-10 with alternative approaches to basic local alignment.

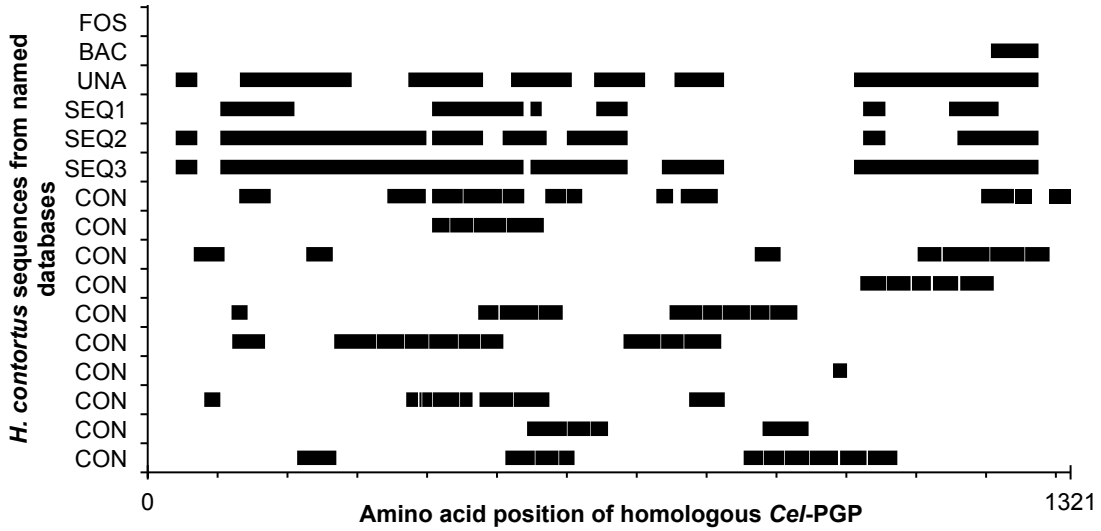


Figure 2: Alignments of *H. contortus* Sequences from Multiple Databases with Homologous *Cel*-PGP.

Residue positions of homologous *Cel*-PGP along *x* axis. Bars: BLAST alignments for sequence from *H. contortus* genomic databases to any *Cel*-PGP. No alignments from FOS database. BAC database provided 1 sequence homologous to *Cel*-PGP. UNA database provided fragmented but not overlapping sequences with no evidence for multiple genes. SEQ1, SEQ2 and SEQ3 databases provided fragmented short coverage but less fragmented longer sequences with each database release. SEQ databases provided no overlapping sequences and hence no evidence for more than 1 gene. CON database provided fragmented alignments with some overlap providing evidence for 10 genes. Marked fragmentation of genomic sequencing is evident in these long genes. Compare to greater alignments with the later databases C09 in Figure 9 and S09 in Figure 11.

Evidence for the number of *pgp* genes from the various databases is summarised below in Table 5. The FOS database provides no evidence of *pgp*, in the next five releases there is coverage but no overlapping sequences and therefore no more than one gene is demonstrated. The CON database contains 10 overlapping *pgp* sequences.

Table 5: Evidence for the Number of *pgp* Genes in Different Databases.

Database	Number of Genes
FOS	0
BAC	1
UNA	1
SEQ1	1
SEQ2	1
SEQ3	1
CON	**10**
Maximum	**10**

3.2 Alignments to Specific Domains

CD-Search reveals which conserved domains are present in the query and enables consistent comparison and positioning of different sequences from both *H. contortus* and *C. elegans*. Figure 3 illustrates the matches from the CD-Search on one *H. contortus contig*. Characteristic for the ABC within P-gps is the ATP binding site and ABC transporter signature. The entire ABC is recognised by the conserved domain (cd) *cd03249*. The sequence also registers with an ABC membrane that is registered as *pfam00664* and represents the six transmembrane helices of ABC transporters (the TMD). Under the multi-domain hits is the pattern registered in the clusters of orthologous groups (COG) database as *COG1132* for both the TMD and ABC. The other interesting hit is also a multi-domain model named *PTZ00265* and derives from the multidrug resistance (MDR) protein in *Plasmodium*. There are a great number of motif and domain matches not included in Figure 3.

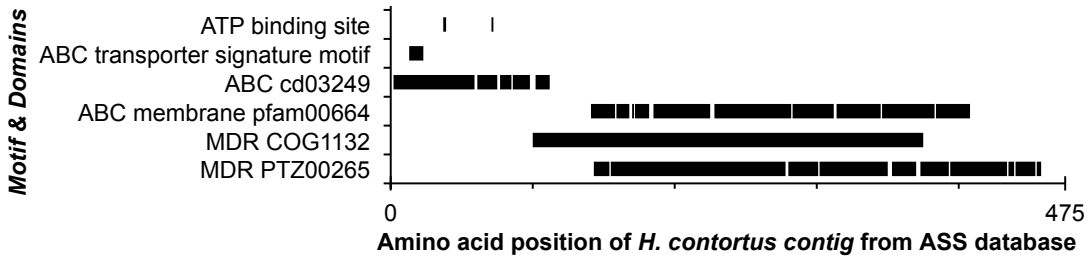


Figure 3: Motif and Domain Alignments from CD-Search to a *H. contortus contig* from ASS Database.

Residue position of *H. contortus contig* 0009203 along *x* axis. Motif matches include the ATP binding site and ATP transporter signature. Domain matches include multiple ABC domain models including *cd03249*. Superfamily matches include the ABC membrane *pfam00664*. Multi-domain matches include the TMD and ABC model *COG1132* and the MDR pattern *PTZ00265*. Matches indicate strong signal for ABC in *contig* 0009203.

3.2.1 Alignments to ATP Binding Cassettes

The ABC model (cd03249) is used to match *contig*s to this domain in Figures 4-8. The ABC (cd03249), TMD (pfam00664) and multi-domain models (COG1132 and PTZ00625) are compared in Figure 9 and then COG1132 is used to match *contig*s to both the TMD and ABC in Figure 10.

Domain Position	Consensus Identity	Query Identity	Query Position	Score
131	T	T	5	**+7**
132	L	K	6	**-1**
133	V	I	7	**+2**
134	G	G	8	**+7**
135	E	E	9	**+4**
136	R	G	10	**0**
137	G	G	11	**+6**
138	S	V	12	**-1**
139	Q	Q	13	**+6**
140	L	L	14	**+6**

Figure 4: Alignment of a *H. contortus contig* from ASS database with PSSM of ABC model.

Residue position of ABC domain model *cd03249* in first column. Consensus residue identities of *cd03249* in second column. Residue identities of *contig* 0009203 in third column. Residue position of *H. contortus contig* 0009203 in fourth column. Score from PSS Matrix of *cd03249* for the residue in *contig* 0009203 in fifth column. Positive scoring for this portion of the domain model just prior to the ABC transporter motif (shown in Figure 5) indicates strong signal for ABC in *contig* 0009203.

The *H. contortus* sequence is aligned by RPS-BLAST to the PSSM in Figure 4 and the high scores for this alignment in the fifth column is evidence for the *contig* covering this part of the ABC.

Domain Position	Consensus Identity	Query Identity	Query Position	Score
140	L	L	14	+6
141	S	S	15	+6
142	G	G	16	+7
143	G	G	17	+7
144	Q	Q	18	+8
145	K	K	19	+7
146	Q	Q	20	+8
147	R	R	21	+8
148	I	V	22	+2
149	A	A	23	+6

Figure 5: Alignment of a *H. contortus contig* from ASS database with ABC Transporter Motif in ABC Domain Model.

Residue position of ABC transporter motif in domain model *cd03249* in first column. Consensus residue identities of *cd03249* in second column. Residue identities of *contig* 0009203 in third column. Residue position of *H. contortus contig* 0009203 in fourth column. Score from PSS Matrix of *cd03249* for the residue in *contig* 0009203 in fifth column. Exact match to consensus sequence at 9 of the 10 residues indicates strong signal for ABC transporter in *contig* 0009203.

Whilst a section of the alignment is shown in Figure 4, a more specific view is shown in Figure 5 with only the residues comprising the signature motif of the ABC transporter. The identity of this *contig* as an ABC transporter is demonstrated by the very high conservation to the signature motif in this figure.

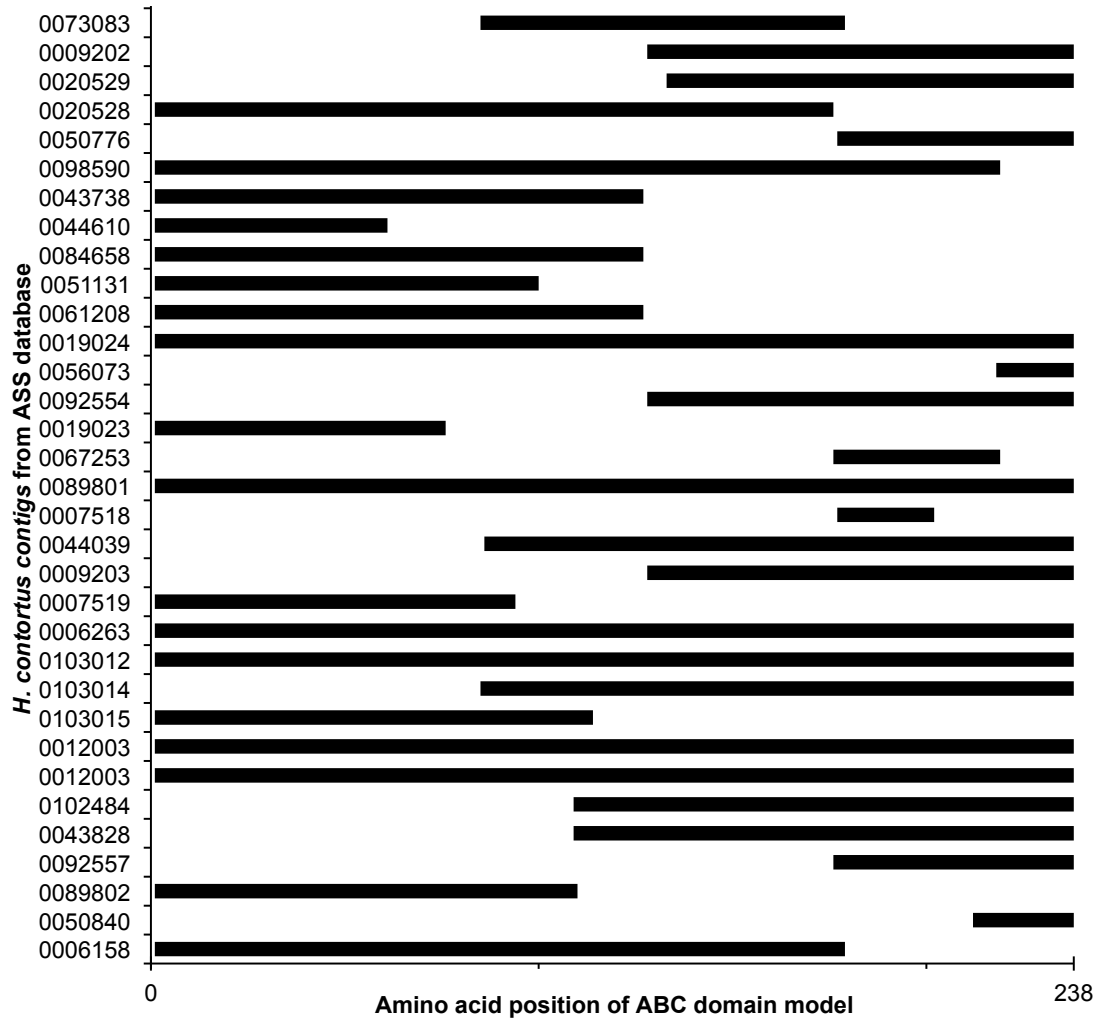


Figure 6: Alignments of *H. contortus contig*s from ASS database with ABC Domain Model.

Residue position of ABC domain model *cd03249* along *x* axis. Full-length of domain model is 238 residues. Fragmented coverage is evident with only 6 *contigs* covering entire ABC. Alignments for 6 *contigs* start or end at the same residue position of approximately 120 in the domain model, likely to be a splice site conserved across genes.

The alignments generated between the ABC and the *H. contortus contig*s are plotted in Figure 6. Of the 62 *contig*s that were most similar to *pgp* sequences in a reciprocal BLAST, 33 aligned with the ABC. Only six *contig*s were aligned with the entire length of the domain and the remaining 27 *contig*s matched part of the domain. The number of overlapping sequences is clearest at the 5' and 3' ends where 18 *contig*s overlap, providing evidence for no more than nine *pgp* in *H. contortus*.

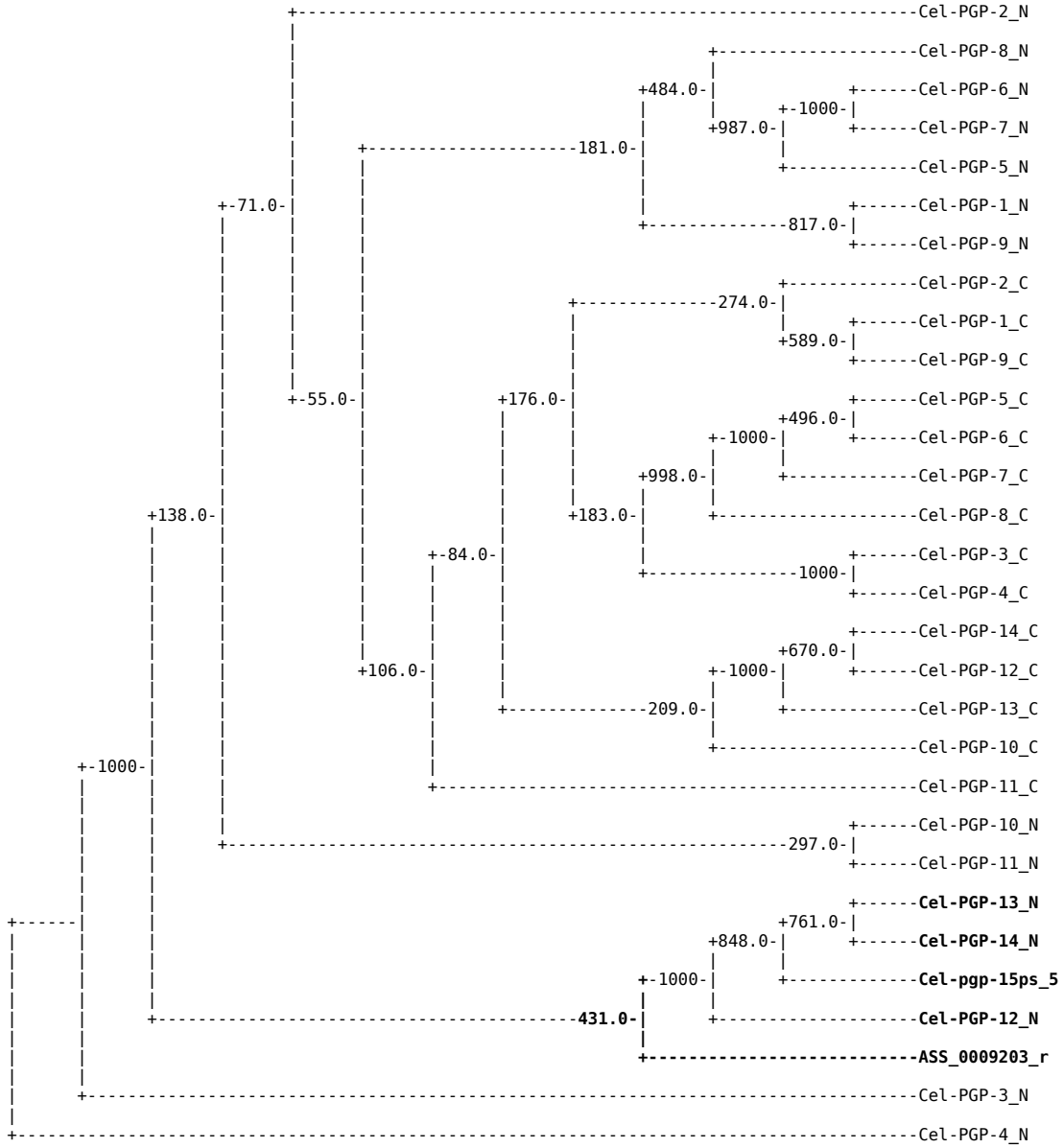


Figure 7: Neighbour-Joining Tree of a *H. contortus contig* from ASS Database with the Aligned Portion of ABC Domain from *Cel*-PGP.

ASS_0009203_r is the complementary sequence (r: reverse) of *contig* 0009203 from the ASS database. N: N-terminus ABC domain. C: C-terminus ABC domain. *Cel-pgp-15ps_5*: 5' ABC domain of pseudo gene *Cel-pgp-15ps*. Numbers: number of trees of *1 000* replicates that divide the sequence into that same division. A ratio of *431* trees of *1 000* for homology of *contig* 0009203 with the N-terminus of the subgroup *Cel*-PGP-12, 13, 14 and 5' *Cel-pgp-15ps*. This is an unrooted tree.

The 33 H. contortus sequences from the ASS database with sequence similarity to the ABC domain were aligned with the ABC and this domain was used as a common positioning system. Both 5' and 3' domains of all Cel-PGP and one individual H. contortus translated sequence were aligned and a phylogenetic analysis carried out with the result of a neighbour-joining tree as in Figure 7. The numbers at each branch represents the number of times out of *1 000* bootstrapped repetitions this separation of the tree was generated. The H. contortus sequence in Figure 7 clusters with a particular C. elegans set of domains (the N-terminal ABC of the subgroup Cel-PGP-12 to 15) in *431* of the *1 000* bootstrapped repetitions, which is very low and does not confirm this homology. However, this contig only aligned with 110 amino acids of the ABC

and hence any further analysis is hampered. When the contig that putatively covers the remainder of the ABC is included in the alignment the result rises to 93%. The phylogenetics assists in producing

Figure 8.

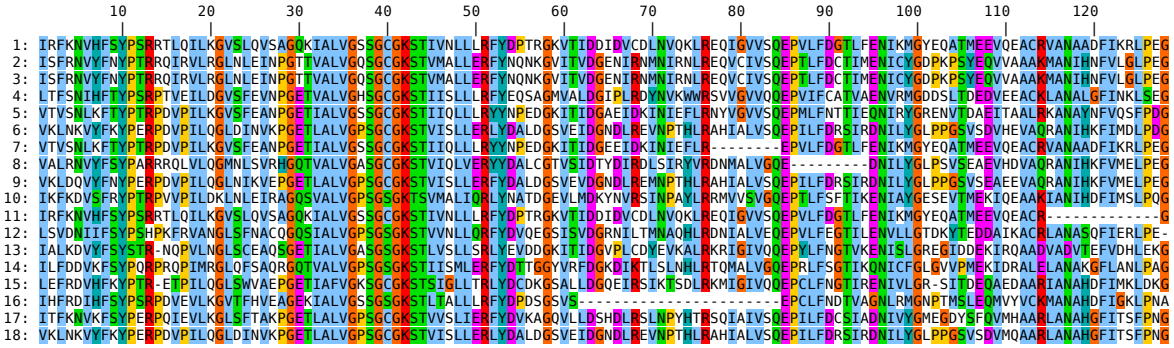


Figure 8: Alignment of *H. contortus contig*s from ASS Database Aligned with 5' portion of ABC Domain Model.

Residue position of ABC domain model *cd03249* along the top. Hyphens: gaps between *contigs*. Letters separated by hyphens: *contigs* that have been edited by the author to match the ABC with another *contig*. Other *contigs* have been matched by the author with no exonic gap and hence no hyphens. Dual number in identity key below: *contigs* combined by author. *Reverse* in identity key below: complementary sequence of *contig*. Clustal-X colour coding highlights residue positions of physico-chemical conservation. Amino acid conservation in ABC domain demonstrated by colour conservation. 1: 0012003 N-terminus. 2: 0012003 C-terminus. 3: 0103012 reverse. 4: 0006263. 5: 0089801. 6: 0019024. 7: 0019023 & 0103014. 8: 0007519 reverse & 0044039. 9: 0051131 reverse & 0102484. 10: 0089802 & 0043828. 11: 0061208 reverse & 0009202 reverse. 12: 0043738 reverse & 0092554. 13: 0103015 reverse & 0009203 reverse. 14: 0084658 & 0020529. 15: 0006158 reverse & 0050776 reverse. 16: 0020528 & 0092557. 17: 0098590 & 0050840 reverse. 18: 0044610 reverse & 0073083 & 0067253 reverse & 0056073 reverse.

Figure 8 is the end result of the *contig*s from the ASS database being aligned with the ABC. Examining how the *contig*s could possibly fit together and then taking account of the phylogenetic support for these combinations leads to these putative 18 sequences of the ABC in *H. contortus*. There are nine pairs and so is evidence for the presence of nine *pgp* in *H. contortus* as opposed to the earlier evidence for ten *pgp* as in Figure 2. The short gaps are between the end of one *contig* and the start of another. Other *contig*s are joined end-to-end with no gaps due to the break occurring within an intron. One *contig* covers a pair of ABC, this is *0012003* and includes the entire sequence of the previously discovered *Hco-pgp-2.1.A* that is renamed by the author from *Hco-pgp-A*. Four *contig*s cover an entire ABC domain each, while 11 domains are covered by a pair of *contig*s, some with exonic

gaps and some without and finally one domain is represented by four *contig*s. The methodology represented by the previous figures resulted in as great a coverage of the ABC domains as possible given the extent of coverage of the genome in the ASS database.

3.2.1.1 Alignments to Both ATP Binding Cassettes and Transmembrane Domains

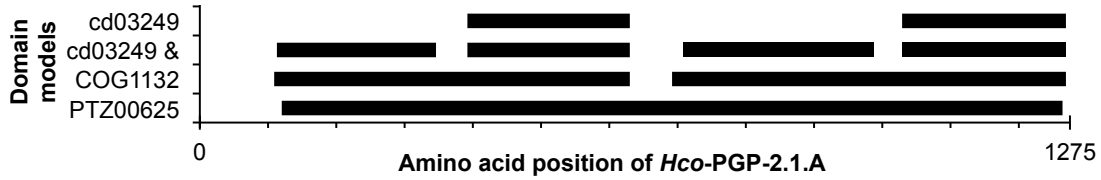


Figure 9: Coverage of *Hco*-PGP-2.1.A to ABC and TMD Domain Models.

Residue position of full-length *Hco*-PGP-2.1.A along x axis. Numbers in graph: length of alignments or gaps for this *H. contortus* sequence to the domain model. *cd03249*: domain model of ABC. *pfam00664*: domain model of TMD. *COG1132*: domain model of single TMD and ABC. *PTZ00625*: domain model of duplicated TMD and ABC. Successful alignment of each domain model to *Hco*-PGP-2.1.A is shown by bars of colour.

To extend the alignment of *contig*s beyond the ABC other domain models are compared. The coverage of these different models through their PSSM is shown in Figure 9. The longer coverage in *PTZ00625* initially appears more useful, however it is based on *Plasmodium* and is not sensitive enough to pick up both 5' and 3' domains but instead disregards one domain and provides a shorter alignment of the *contig*s than *COG1132*. This *COG1132* has as high a sensitivity to *H. contortus* sequences as the single domain matrices of *cd03249* and *pfam00664* whilst improving the coverage of the *pgp* and so is used to generate the alignments in Figure 10.

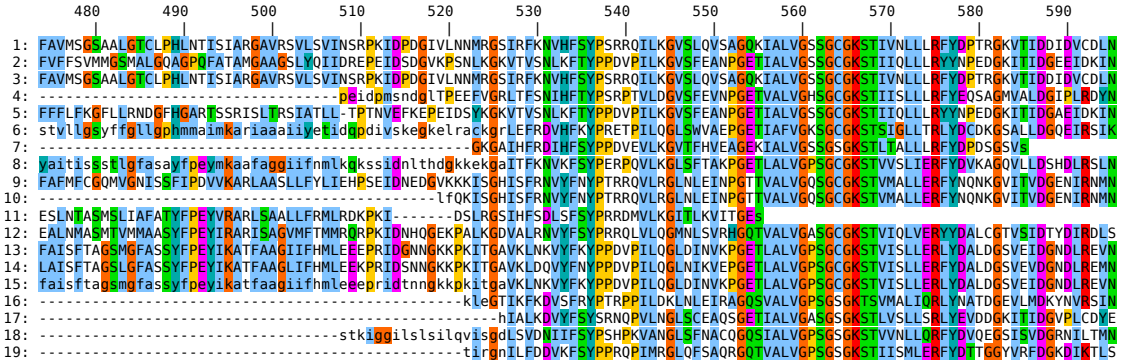


Figure 10: Alignment of *H. contortus contig*s from ASS Database at the Intersection of TMD and ABC Domains.

3' end of the TMD is on left; 5' start of the ABC is on right. Clustal-X colour coding highlights residue positions of physico-chemical conservation. Greater conservation in the ABC indicated by the colour conservation on the right. Upper case letters: BLAST alignment. Lower case letters: edited alignment by author. Residue position in *COG1132* TMD & ABC domain model along top. 1: 0012003 N-terminus. 2: 0019023. 3: 0103015 reverse. 4: 0006263. 5: 0089801. 6: 0098590. 7: 0044610 reverse. 8: 0051131 reverse. 9: 0012003 C-terminus. 10: 0103012 reverse. 11: 0009203 reverse. 12: 0007519 reverse. 13: 0019024. 14: 0061208 reverse. 15: 0089802. 16: 0043738 reverse. 17: 0006158 reverse. 18: 0084658. 19: 0020528.

Of the 62 *contig*s that aligned with *Cel-pgp* on a reciprocal BLAST, 33 covered the ABC and a total of 53 are aligned within the COG1132 domain model incorporating both the TMD and ABC. Those *contig*s aligned within the COG1132 domain at the intersection of the TMD and ABC are shown in Figure 10. The alignment clearly shows a greater number of alignments on the right (18) than on the left (12) and this number decreases further to the 5' end of the TMD, beyond the region shown in Figure 10 where the lower conservation decreases the sensitivity of the method to detecting *contig*s that cover these regions as opposed to the ABC. In particular, sequences that only covered the TMD were not efficiently detected.

Table 6: Evidence for the Number of pgp Genes in ASS Database.

Domain	Number of Domains	Number of Bi-domain Genes
ABC	**18**	9
TMD	**12**	6
Maximum	**18**	**9**

The evidence for the number of *pgp* genes is summarised in Table 6. The ABC domain is present 18 times in the database and hence is evidence for 9 genes,

whereas the TMD is only found in 12 sequences and does not support more than 6 genes.

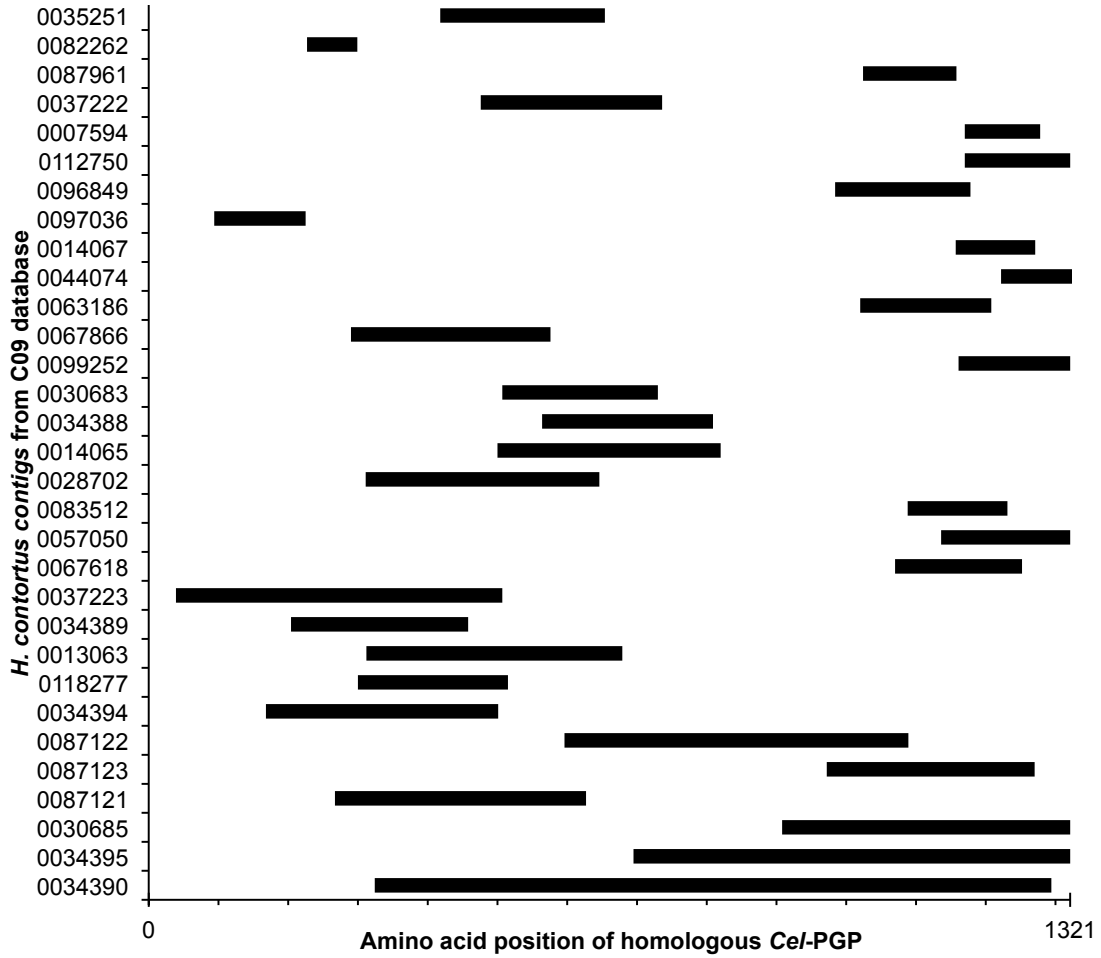


Figure 11: Alignments of *H. contortus contig*s from C09 Database with Homologous *Cel*-PGP.

Residue position of homologous *Cel*-PGP along *x* axis. Putative ABC signal is visible from the number of alignments around residue position 550 and 1150. Majority of *contigs* are relatively short with larger gaps; only one *contig* putatively covers both ABC.

The *contig*s that were established as homologues to *pgp* by reciprocal BLAST are plotted in Figure 11 showing the regions of alignment. There is varying evidence for the number of *pgp* genes over the length of the putative genes from nil start signals to 12 close to the 3' end as summarised in Table 7.

Table 7: Evidence for the Number of *pgp* Genes in C09 Database.

Position	**Number of Genes**
Start	0
550	10
1100	9
1200	**12**
Maximum	**12**

The C09 *contig*s are generally longer than those in the ASS database but are still too short for *pgp* full-length alignment therefore the next stage was the use of *supercontigs* as in the figure below.

3.3 Alignments along Entire Gene

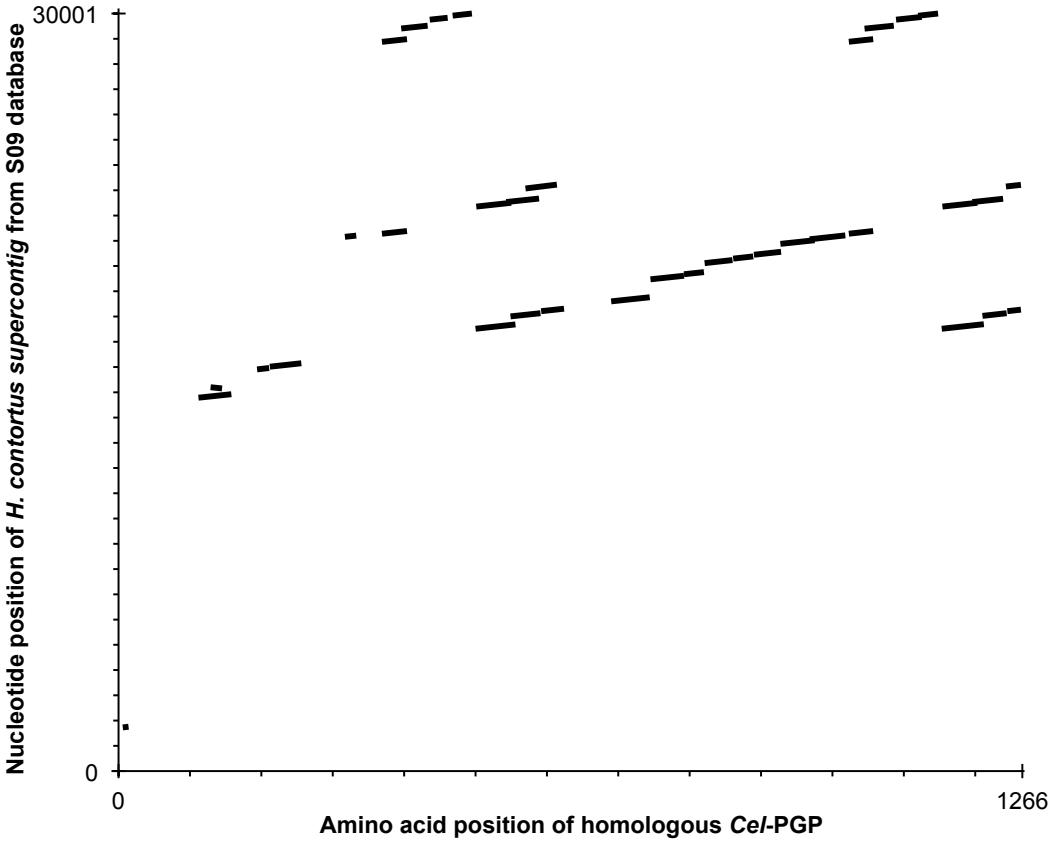


Figure 12: Dot Plot of *H. contortus supercontig* from S09 Database Aligned with *Cel*-PGP-4.

Residue position of *Cel*-PGP-4 along *x* axis and base position of *H. contortus supercontig* 001880 up *y* axis. Dots and lines: regions of identity with putative exons separated by vertical jumps up the *y* axis according to putative intron length. Direction from bottom left to top right indicates agreement to the provided sequence rather than complementary. Pattern suggests alignment from residue 100 to the stop signal from base 15 000 to 24 000 approximately. Characteristic 4-cornered box of bi-domain signal due to 5' *H. contortus* putative ABC domain matching both 5' and 3' *C. elegans* ABC domain and the 3' *H. contortus* putative ABC domain matching both 5' and 3' *C. elegans* ABC domain.

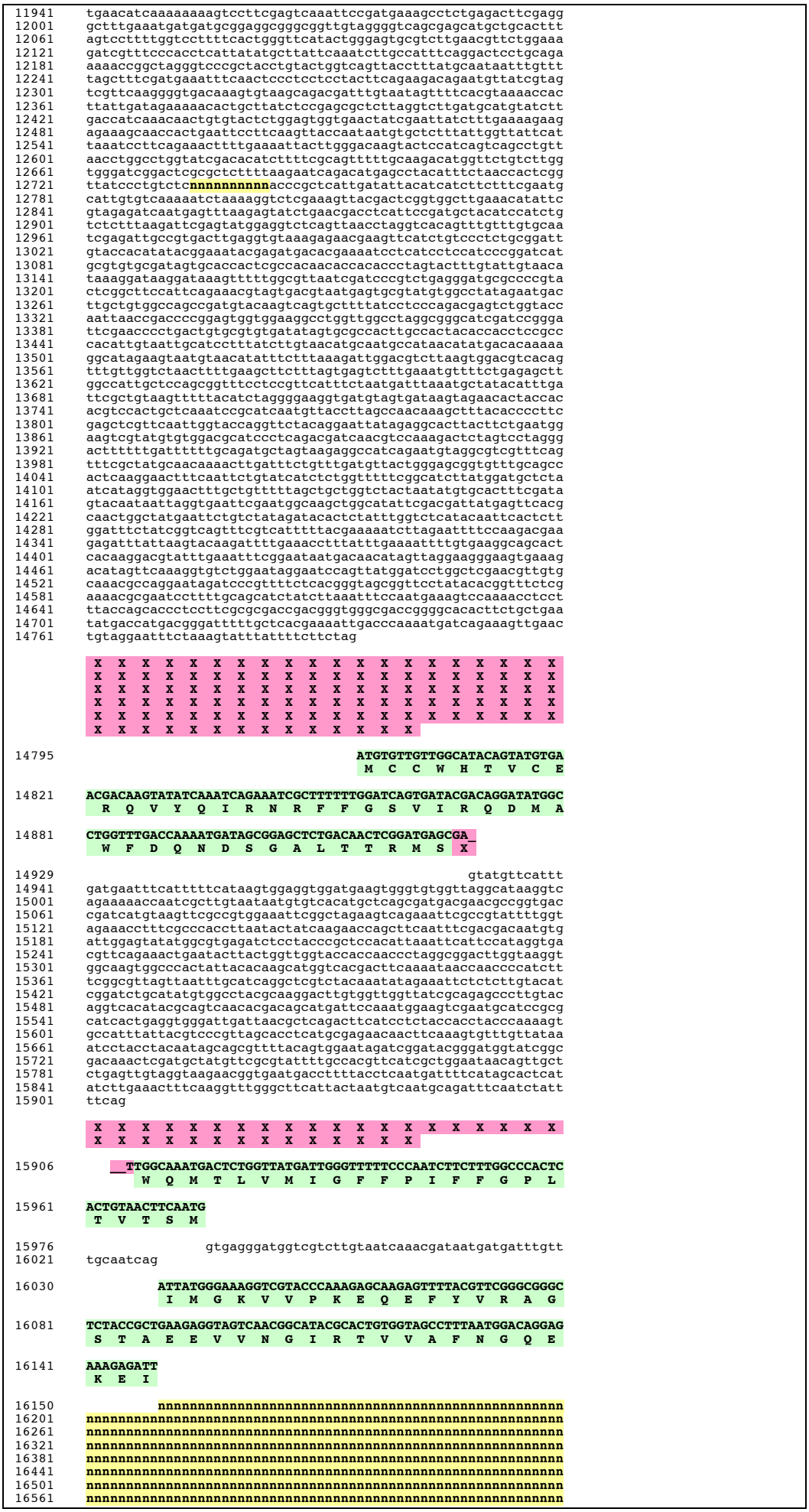
The HSPS generated by BLAST between a *H. contortus supercontig* and the homologous *Cel*-PGP can be plotted as in Figure 12. The conservation of the 5' and 3' ABC is apparent in the characteristic mirrored HSPS, resulting in the appearance of four matches from two genuine matches because the N-terminal ABC of the *C. elegans pgp* is aligned with both the 5' and 3' ABC of the *H. contortus* and the same for the C-terminal ABC in *C. elegans*. There is a lack

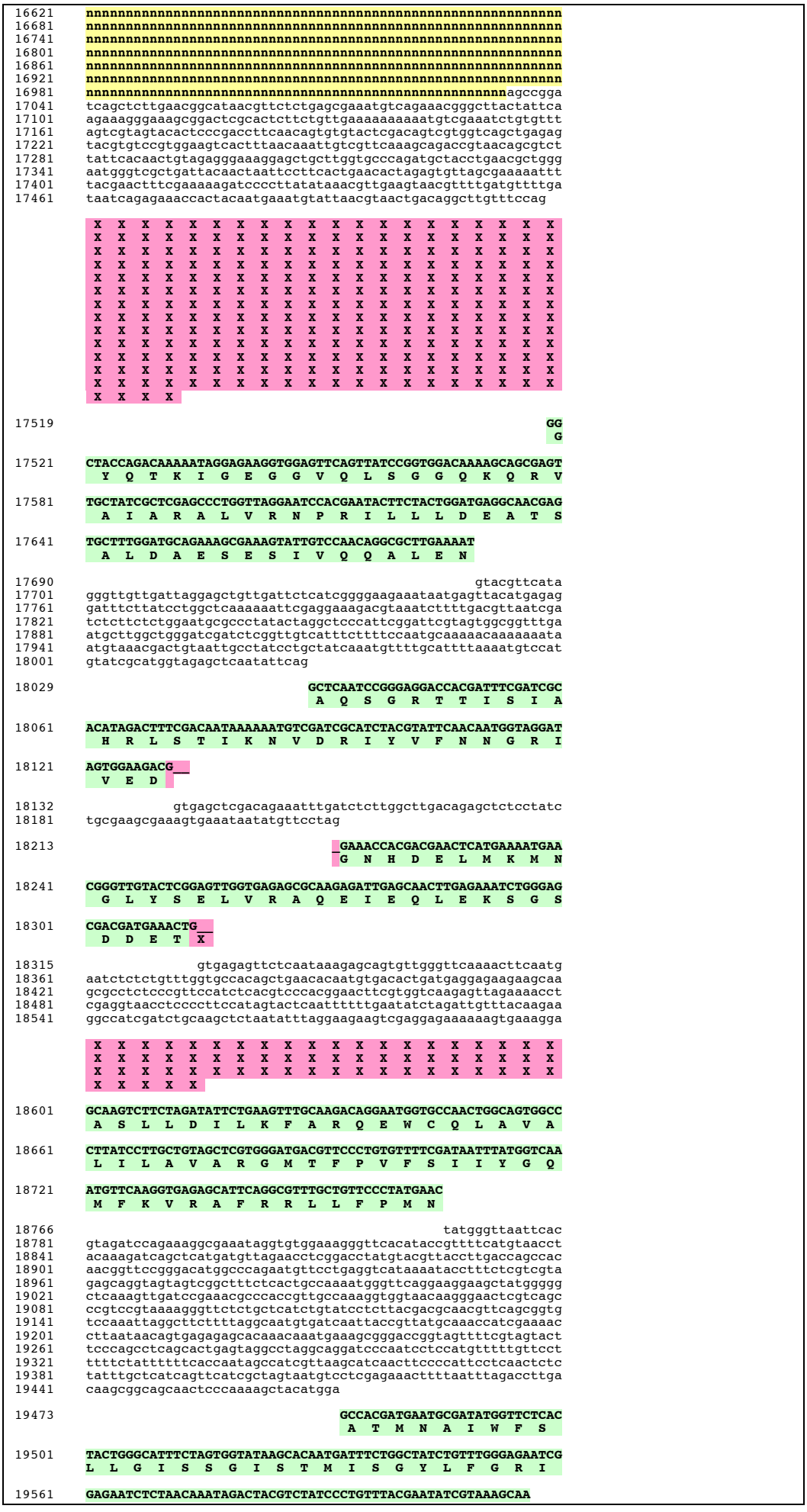
of alignment to the amino-terminus (N-terminus) and in the hypervariable cytoplasmic loop between the pair of domains using these methods. This was repeated for the other 14 *supercontigs* that contained HSPS to *Cel*-PGP. The alignment most likely to be true is the longest diagonal line, in this case it has major interruptions and the long gaps remain even within highly conserved domains suggesting the sensitivity of the alignment software is not at fault. The annotated *supercontig* in Sequence 1 below shows the reason for these gaps.

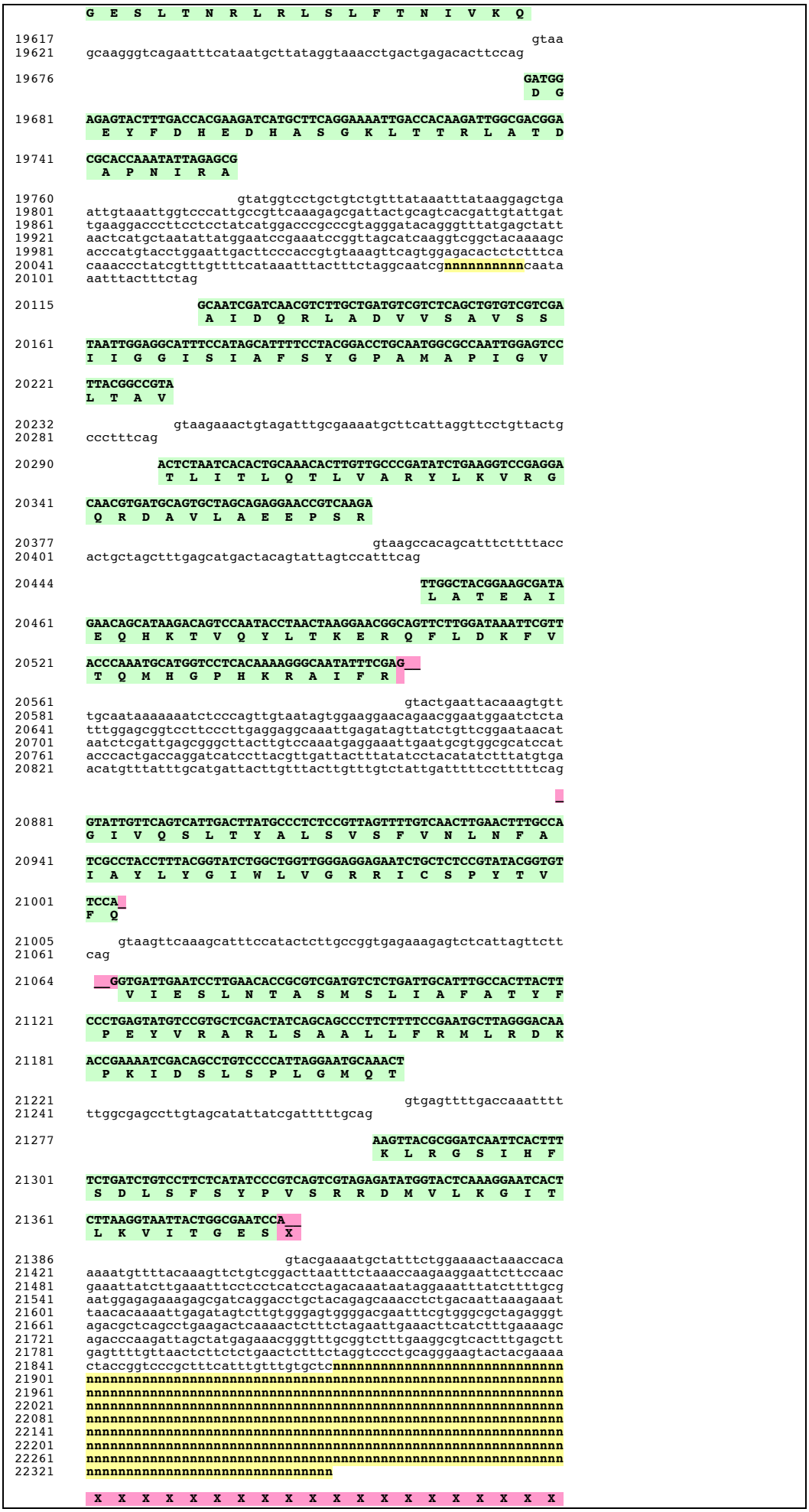
>*Hco*_S09_*supercontig*_0001880_1-30001nt

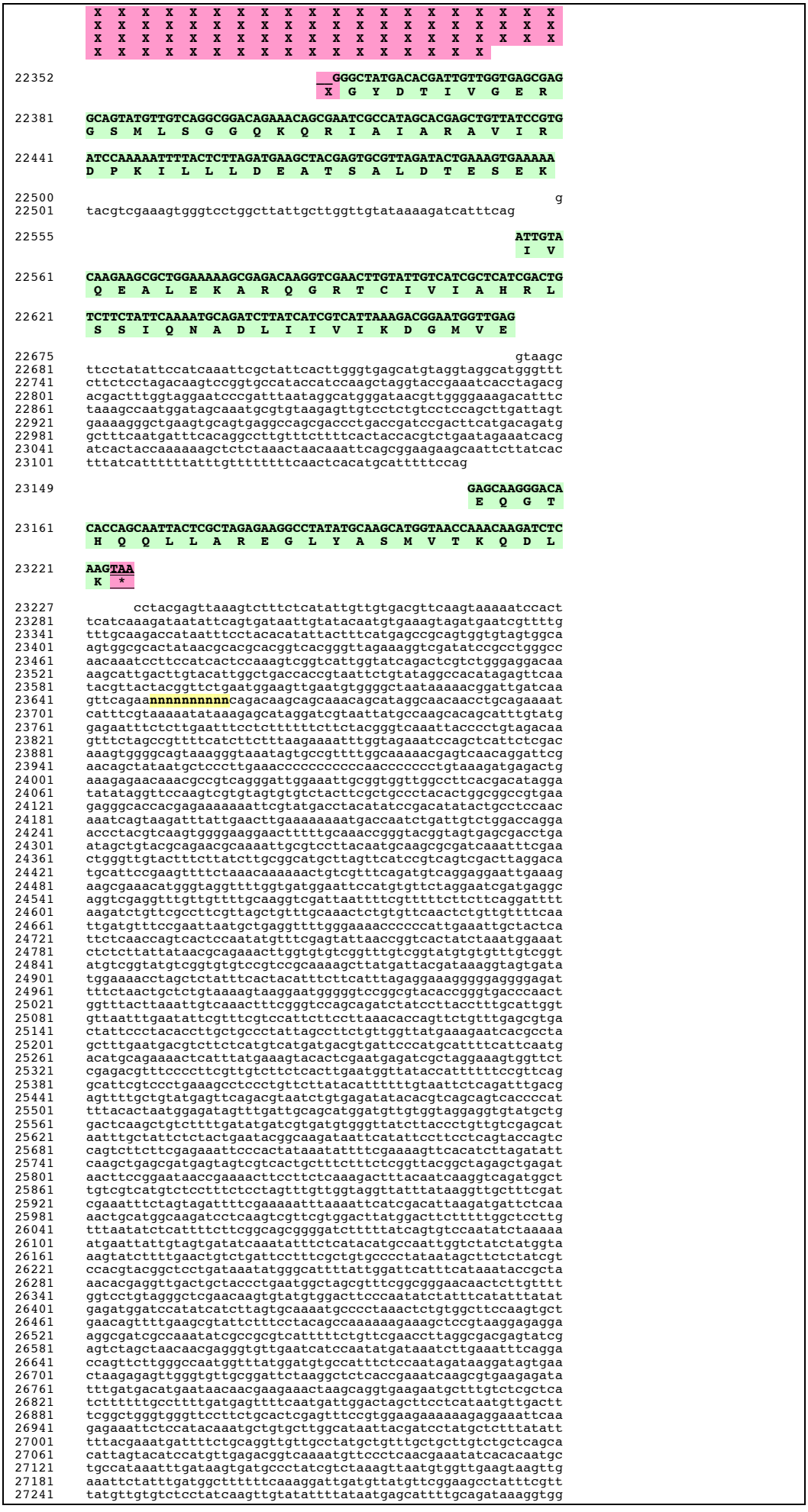
```
1       cctttcctgggagttgattacttggctgtaggcgccaagtgttctccaccgatgattcac
61      caggtcgacgtttccagttcgccacctttcccttcattgtaccagcagaaatcattccta
121     aaaatcaaaataaattcaaaatcaacttttttcctggtaggagggctaaaagccattaat
181     gccaaagtcactagtgacagttagtgatttccgattctgatgatgcgtcactattttgat
241     tcctgtcgaaagttaccggaaatctcctgaggttttgatactgtgctccaccagtaaatc
301     aaggcctgtctaatacaatgagtccagctttttgagctttcctagtacctactaaacaac
361     tgtggaagaagtaaggtgaggttcgccagaatgctctctttgatcctatgcacgtgggaa
421     gtggccctttgccatgaacgcgaacagaacttcgggagaatttcgctgatcagctgtctg
481     tattttgcaaccgaatgttgtttcaaatagtcggtagttcccactgcttacttacttta
541     cgcttagagtaatctacacaaaggatgacatcgtctgcctcgaatccttttgatgactgc
601     aatgagatggtgccatttcttctgtctcaagatcccatactcgtaaatctctttcttgtt
661     ggcaaataagtaagagattttctctgagacagatcgaagtacccccttgccagacaatt
721     ttacctgaaaccatacaagtaagtgattctcttagtttgtcgcctgcttcacagagaaag
781     aacggcatgtctataccttcaatttctcaacagttgtagagtctggcattatagttagat
841     gataaagcatcatatcacttgtcaatgcgatcagtatggctttatggtcatagtaagcga
901     ggaaatcgactcgatagtccagctgaagtaataaaaaataacatcgagaatgataaggca
961     tgaaaggataagagatatttttcggcatcaatatctccccttatggtatgcgccctgtag
1021    tcgttttccatccagggtcgtggtggtgtttgagttcgagtgggacctatgtatgtaatg
1081    acttgcactaatctgggtacaaataagagccttcatcctccctgtggtgaagtgtggtcc
1141    agatatagcagtaaaatactttgcctgccacgcgattgctttgtgtgtcttcttgtgct
1201    tctgtcgtactcctgtgtacttctcactttattgttgacggcgtacggcagcggaatcga
1261    aacagtggagttggacaagaagaagaagaagaagaagaaatcacccttctgtgtgtctta
1321    tatactatagttgattagatccacataaatcttactatcataacctcatttttacatgac
1381    tttctacacgcaacggtaatgtttcgtacactcccatacgggtctggcctacaccagagg
1441    gatggaacagtttgttagtcttagtggtatcgatctgggattcgatctcggagccatgat
1501    tgtaagaatgagccttttgacactacaccaccccaatccattccaaaaatatagaacaaa
1561    gcagaaaagtgcgttcagtgttcaacacggacctcaaattgagctgaattttctgtccca
1621    cttcaattttctctgacaaattcgaggaatagtagcttgtagaaaataaggcgtagcgtg
1681    agttccttcccagttttccaatccccttgaatttctacatatgaatttcaacacgtggaa
1741    cttggagggaggccatcatcaatcatgccacaagtacagtaaattgaggggatttttcaa
1801    ggtcgacattacgtggctcaatctgatactgctagcgttttgatcaatcacataaaccac
1861    tcctccttcggtaccaacagcaaacgttgttgtgagcgagatggacatttgagctgtatt
1921    cacctacaaaggggaaacgatgtattaacaacttcagcggaggggaaactttggaagcct
1981    tctttagcccttcctaccgtttcctctttcgttgaagttccattcgtcgcctgattagct
2041    tctccaccaaggtgtgcaagttcattcctaaacactctaacgggtaagtgtatcgatagg
2101    aacaacattataagtcccgcccataaaaaaagccttgagctccaagccaggtttctgaag
2161    atcttttcctgggatttcctgataccatttctgcgttggctaaccttcccataactattc
2221    tgaattcataactattctacaggcggaggtggtgtagtggcaagtggcgcactatatcac
2281    acgcacggtcaggggttcgaatctcgaatcgatgcccgcctaggccaacctagtcttcca
2341    tccctccggggtcggtaaattggtaccagactcgtctgggaggatagaagcgctggcttg
2401    tacatcggctggccaccgcaagtcattgtataggccacatacgcactcattacgccacta
2461    cgatttctgaatggaagtcgagtgcggggcgcatcccctcagacgggattgattaacgccaa
2521    aaactttacttaattttaactattctagaatataggacggaaaagatatatcgacaaaaa
2581    aggcgctatggcccggaaatcaaacccatagatgacacacgccctggccaaccaaaagta
2641    gcgcttaggagtcgaactggtattggacttgtctcggaggataaaagccaaggccaaaac
2701    gataacctagcctttgtacgagtcagtgttttatgctcccagacgaatgtggtacgggag
2761    aggtgggaggcttggtcgtcctaggcgaatgtggatgtgggattcaaactcatgaccgtg
2821    cgtatgaggtagttccgctttccactacactatctcctccctgaagatgaagtcagaat
2881    agcttcccgatctgttgcgatcctatagatgccgatgaatgaccaacttacgccttcact
2941    cttctgcaagctgctgtcggtttgtcacccacactgcattcaccttctttttctgttgag
3001    cttttggtgaccagactgtaggcgatgcattttccatcctgcaaattaccaaaaattctc
3061    agctggagctgtacatatcactagagattagacatctttacctccataaccacccaaaga
3121    gacttgccatcggcgctccactgtagaaaggtgaccttacttttttagttcctcgtctacg
3181    ttgaactctagaatatttgaaagtcattaacactcgcgtcttcagagcttccgagggttc
3241    agaggtctccagaagcagtcggtaaaaactccaacgattgtttgttggatagaaccgttc
3301    aaaacatgcttcactctatttcatagtgaagttagatgaaattatatctcaaatacatcc
3361    agtcgggtaagaaacagcaagaaacctacctatgttggcctccggtgataggattgctag
3421    caatccatcggtccaagcaatcgctacaagaatctctaacggattccacgaaagggctga
3481    gacagtgctcttccgttgtaccacaacatcttcaagacgaacacccttttaatttcaatga
3541    tcaccaagtttatgagttcattaagttcagattcagaaggcagtagcatcgactccgttc
3601    tctgcttactgtcagagaattgtaatttgtagcagagttgggatagttgccagatgaatc
3661    agtttccctgtctcaggagaacatgtctattatggccgtgataaggattctgcgttatac
3721    acacttttccaagggaagcattacttccaggaagaatagtattgagtaatggagggtag
3781    aaagatatggatgaatttcgcaaagtaccgtaatgaaccaatttgttctgctccctctgc
3841    aatttttttttctcctaaaaggtttattttgcaaacaacttgctccacgagagatcgtgac
3901    gaacgtgcaggacgtcataggttgtctggggcccacaagggtaatgacggtgccccttcg
3961    ggttcatccttttgctgaaagaccagatgcgtttgcacattttctttgaagtataaaacc
4021    tgttagttcggtgaacctatgtccaaagtcccgatcccaaagtacttctttcgactaccg
4081    acgggtatgttagagtgcttagtcacccccctagctccgtaagattgtagaataggggtgt
4141    cccttatgtaaacggttggatatgcgggtctggacagtgatccgctcaagctccatgctg
4201    gtagttgcataaaaaagacccaaagacccacatttacaaaataaagtataattcataatg
4261    gttaagcccattatttcattaatcatgaataaaggtaagaaataacaaaactacacacct
4321    tatgagtaaagaagttaatctgaacagaggactgaacttgagaagccaaacacagccagc
4381    cactgcaaggagaccatgcaaagatattcttgtccatagctaactcctcagaatatcatg
4441    agtgggtcaggaatcatcagggccatagaaaacccgtgctcgaattcatgccgttattcg
4501    gtacaccgttgtcatgacgaccgaaaggcagcgcacctgcctaatgtctgaatgaggcgc
4561    tcctactgctttttgagggtcgttgaaataaacaggatcatataagcacaattcaggttc
4621    tatttcagagtgtcgcaaagcagtcttacagcttctaaggaaatttcgagaactaggaac
4681    tccagtaccactgggcgagtcattagagacatcgagacagatcgcctccccccattacca
4741    atccctcgacctcattttgaagctccaaatacactgattactcaatagtctattcacaaa
4801    atacctccattgagaaccatgctgtcccctgaaatacgaaaagcaaacagatactttact
4861    ctcaccaacaacaccacacaataaataaacaattcaataattatcgatacgtcaataacg
4921    ctttcctcacaagtacgtcctataacagcgtcctaaacggcaagtcgaacttaattttca
4981    agtttaggcctctatattttctatataatagggttaaatacaggctgagaatggattgcg
5041    caaaccaacaacagtaccactaccccgtcctatgtcgatcacaaacaacattaaagtaag
5101    atcacagatcgtaagaaattgagcccacgagagaactagtcaatttttcttcgatgccctt
```

```
5161    ttcctcgctctcgccttttcagatggagaacccggaaggcactgctttgacaggttttcc
5221    atcaacaacctccgcagtggacttgatttcagccttcttctcttcctgatnnnnnnnnnn
5281    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5341    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5401    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5461    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5521    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5581    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5641    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
5701    nnnnnnnnnnnnnnnnnnnatggtatgaaatcgtattttgccaagcggaacaggaaggta
5761    cttgactgtgatgttcatatgagtatctccttttttcatctgaaataaaggaaattattt
5821    ccagcataagtaaccgtgaggggacaactttgcatgtataggatacttagtaaaacatga
5881    aaaatgaagtatgctgaataaaaggagctatgattaagctacttgccaggctacccatga
5941    agtgaagtttaactttaaaagcaacataattactggtgaaaactttcgaataagaaatca
6001    tgcaagaaaccctaattgaacgaccggctaacacacaactaacctcaaccagatctttct
6061    cacgactcagcaaatcggcaacatagagtagcgggaaataaccctcatcatttccattca
6121    atctggaatgttgaagggtggcaataacggctttgaaagttagatgttaaattatatatg
6181    gagcgataaatcttacgagatttccggtaactctggaagggacaaatcgggagagtcttc
6241    caaactcattacattcactacagatcgaatatgagtgataggaggaagtttctgaaataa
6301    ataatgatagacgagccgtttccaactgcagcaactcagcagacaggacttttcgtcacc
6361    ttgggctacattgagacctgattacacttccctactcaactactcggttgactacggcat
6421    ataaatggacagaatgctactagttcccattaaagtcctttcgctctgataagacaatgg
6481    acacatttaacagtcctgaacagcaggcaatccacaacttcacagagggaaaaggcaaat
6541    tcctagcattgggatttactgggtgacttccacggttctcgcagctccatcaactgccag
6601    taccaaatagttcatgattttcgataagccgtcaacattggaacttcacttacacaggcc
6661    tacgagatttgctgaacccaagagggatttgaaaacgaactttaaagaaaattcaaaacc
6721    ctccacgaacaaggtcatctttagcccgaatttaggtgttgtgacaatttaatatcccca
6781    ggataggcgacctttctcggtaattttcgctgaattgctgcgggcgttttcaaacagcat
6841    gcccgcacaaggtcatctaactagaatctaaacatagattcctgggcgaaaatcacgacc
6901    catttggcacgtgaacatcatgaacaaccataacaaatcacaaaccgtgcgcgattaatc
6961    cactaataaagacaaaaaagcaaaaatactgacttctctcagttcgccatttatcaagtt
7021    gaatgtacgttcaggctcacgatatgcacttgcttgaatcccttgcagagtataccatgg
7081    agccgcctaaagatcgtaatagcgctgttagcttgacagtatgattaaagatgcaggtgt
7141    ggatttaacggacaaaaattaccatgtggtttacattttcaggcgtctcgttgggaacta
7201    gaacgaagaataagaacatgctgccgttttgtcgagtcgtaagaggaagggatactttca
7261    cctgaaacaatcctataaatgaaaaataagagaatctagaccgatagccgataaacaaaa
7321    aatgcaacgaaagtccatttgagcctcagaacgaatcaatgcaacaagacagttcggaaa
7381    tatttgcgaagtgttgtgaccttatttgctgcatcccaattggcagctttcgctggcact
7441    tgctattcttaccatattcatctaatcagatcaatcatcttcacccctttgccttatta
7501    cccatcaaaactttgaaacaacttcaccttcgttcgggtctgggggttgttcgagggtgc
7561    caagttgcactactgctaaacgtaggtccgaagatcaactcgatacacgctcatggaaac
7621    aatggcaattcaatggcaacagactaatcccggagcactaaattgaacggttacgtttcc
7681    gctcccccacaagtttggtaagccaaaatctttttcgtnnnnnnnnnnnnnnnnnnnnnn
7741    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
7801    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
7861    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
7921    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnataaaggtaaataataaatacgatgtcatc
7981    actgaaaattgactttcatgaaaattttccaaaaaatgttcaaaaaaaccaccgaccgga
8041    aacaagcaacgcgatgagcagttggcatgactattataaatgcgaagagatcatcgatga
8101    acgtattgaaagcctgaagaaaaccagtagagagtgattttgggaagcatttaaaaatat
8161    gtagatcggttttaaatgggtgaaaaaacaaaatgagcagaacaatgaaggtgaaagact
8221    gtaccttgtacgtaaaagctctccatgggagatgcgccacactcttcaatttataattga
8281    caaacaactgaggaagcatgaacaagaagccaaacgcgtagactccattagccattgact
8341    caagcacccagctataccaggacttatgtggaacatacatcaagctgaaatcaagacctt
8401    tcttttaggttttttcatcgcgatgtgcgtaatccctaccactgcagatgtttattccta
8461    tagaaagccgaacgcactacaatatcccaagctactaagtagggatcggctcggatttgt
8521    agatttcaatgatgcgaattcaattgagaatataaagccactgatatatttacctgtaaa
8581    tggcgcctccaatgcatagtggagtcataagatatgccaaatatttcatagcttcattat
8641    caaacgcatctgtctcttcttcctcagctgaatgctttccgaatttcagcccttgccatg
8701    aaaacgagacctgcaaatcataaaggaaattactacgtacgcgtacacaacctttctgat
8761    ggaaaagagcaaaagaaagaaagcaagcaagaaagcagaatcattcaaaaccacgcctga
8821    cactgatgacatgggttcaacacaatcaatgtcctgattgacaacaacttcacaggcact
8881    aaccttgaaagccttttcagtttccagaattcaatcaaaaccgaaattcccgcaggtac
8941    aatcacgaggagacttgtttcttgatcttggagatagaagaaaattacggtatacgaaaa
9001    acatcgccatagcaaagccctggcagattgaaatgggtcatacggtattggagaagatta
9061    caaataaaaccagtcaatgaagataagtatttcgtgaaacgaacttacttgctagaaagg
9121    ccaaccatagattctcgttctcgccaaaaggatatgtcattcttgaatgacaatacatcg
9181    aaaagaaggtgaattgaagaaacgaacacggttataccaagaaggtaaagactggtgtct
9241    ccaaagatagcctttacctggaaaaaatcttataaaggcaaaacaagatacttgacccca
9301    aattaccgaacagaggaaattttcggttctagcatattactggtgttactgaacaattgg
9361    cagaaaagcgtatacgacagtcgactcaatgaaacttttttaggccctatcaatatgaaa
9421    actccccaaatcatgaagtaggcaaaccgcaacaaaatgtattacaaaaccctcttagat
9481    ctcaggtacagtcatagctttaatgaaaacatannnnnnnnnnaagaacaacatcaagaa
9541    aaaacaaagcgtcgtaggaagaggttgaagaacggatgggcccattctcgtcttggcgtg
9601    tcgcgacacccaatcgcgcgtcgtcgcaggctataaacggggagattttaggatttctgt
9661    ggttctggactggaattctgcagcaactcgagttacagggagaaaaagcgaatgatttag
9721    aacaaataaaattgagaacattcgaaagtttcaaatgagaagtacaactaatgtgggaat
9781    agcaaaagtagtcacgaagggcatgggaagagaagaaaaatcgaggcacattcacaactg
9841    ttaccggtctgatttttcaagaatgcacttatgcattgtttattccataggcctacagtt
9901    aattgtttacgaaatgtggtaaacaataacaaatacaagttatgtgagactagcaagaaa
9961    actcacaatcggcactaggattccagtttggtcctcagtcggtcggtcagctatagcgct
10021   tattgagtacatggaccttcctcctatgagtccctatgactccctgtcgctgcttct
10081   aggacgatgctctgaagataggctgaataaatgcttcgcagcttctaaagctaggactgg
10141   aagttccctaacacttaaagggaatgctggacgtcattttatagaaacaatgaatgattg
10201   acagagctttgttcgcagtactccggcttccctccaaaatcacttttttttcggaaacac
10261   tccactgcttcaattggaacgggtgcatgaagttattcaaacttcttctagccctagagt
10321   caaagtgcttctcttcgtgaagtcgtctgaataattttatgcggtcaccaattcagagcg
10381   aattcagaggattaatccgccttttgtgctgcagttcactaccttaccttgacgcccacc
10441   actaaactacctgtagcttcaaatgaacgccatcgataacaacttctcaagaaacttcca
10501   acatttggtactgcttattgcctatgaaagcgaagttattgaaaagaaaagtttcgggtc
10561   aagaaaagttaaccacagtgggacgtgcactcctactgcccattagaacgctgagagaga
10621   taaggggccccgcacataaaggtgcttaggaatgttccttagccggattcctatagtttt
10681   ctgcacctatacaccgtagaagaggaggaacaccggggagggtaagtcaattgaacggtc
10741   aattcttgcccccatcccagcttcttcagctttttgcacaaggtgagtatctcacatata
10801   ctattgtaaaaattgtgagtaaggtcgaccatctaccctccccctctccccttcactacc
10861   ccctannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
10921   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
10981   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11041   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11101   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11161   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11221   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11281   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11341   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11401   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11461   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11521   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11581   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11641   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11701   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11761   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
11821   nnnnnnnngctatcggagaagtgagttagtgctttccttctttggacatatttccccgat
11881   gttgtcctgaaggagcgattcattcgaatctccaaaatactggttagattagaatgtttt
```






G E S L T N R L R L S L F T N I V K Q
19617 gtaa
19621 gcaagggtcagaatttcataatgcttataggtaaacctgactgagacacttccag
19676 GATGG
D G
19681 AGAGTACTTTGACCACGAAGATCATGCTTCAGGAAAATTGACCACAAGATTGGCGACGGA
E Y F D H E D H A S G K L T T R L A T D
19741 CGCACCAAATATTAGAGCG
A P N I R A
19760 gtatggtcctgctgtctgtttataaatttataaggagctga
19801 attgtaaattggtcccattgccgttcaaagagcgattactgcagtcacgattgtattgat
19861 tgaaggacccttcctcctatcatggacccgcccgtagggatacagggtttatgagctatt
19921 aactcatgctaatattatggaatccgaaatccggttagcatcaaggtcggctacaaaagc
19981 acccatgtacctggaattgacttcccaccgtgtaaagttcagtggagacactctctttca
20041 caaaccctatcgtttgttttcataaatttactttctaggcaatcgnnnnnnnnnncaata
20101 aatttactttctag
20115 GCAATCGATCAACGTCTTGCTGATGTCGTCTCAGCTGTGTCGTCGA
A I D Q R L A D V V S A V S S
20161 TAATTGGAGGCATTTCCATAGCATTTTCCTACGGACCTGCAATGGCGCCAATTGGAGTCC
I I G G I S I A F S Y G P A M A P I G V
20221 TTACGGCCGTA
L T A V
20232 gtaagaaactgtagatttgcgaaaatgcttcattaggttcctgttactg
20281 ccctttcag
20290 ACTCTAATCACACTGCAAACACTTGTTGCCCGATATCTGAAGGTCCGAGGA
T L I T L Q T L V A R Y L K V R G
20341 CAACGTGATGCAGTGCTAGCAGAGGAACCGTCAAGA
Q R D A V L A E E P S R
20377 gtaagccacagcatttcttttacc
20401 actgctagctttgagcatgactacagtattagtccatttcag
20444 TTGGCTACGGAAGCGATA
L A T E A I
20461 GAACAGCATAAGACAGTCCAATACCTAACTAAGGAACGGCAGTTCTTGGATAAATTCGTT
E Q H K T V Q Y L T K E R Q F L D K F V
20521 ACCCAAATGCATGGTCCTCACAAAAGGGCAATATTTCGAG__
T Q M H G P H K R A I F R _
20561 gtactgaattacaaagtgtt
20581 tgcaataaaaaaatctcccagttgtaatagtggaaggaacagaacggaatggaatctcta
20641 tttggagcggtccttcccttgaggaggcaaattgagatagttatctgttcggaataacat
20701 aatctcgattgagcgggcttacttgtccaaatgaggaaattgaatgcgtggcgcatccat
20761 acccactgaccaggatcatccttacgttgattactttatatcctacatatctttatgtga
20821 acatgtttatttgcatgattacttgtttacttgtttgtctattgatttttccttttcag
_
20881 GTATTGTTCAGTCATTGACTTATGCCCTCTCCGTTAGTTTTGTCAACTTGAACTTTGCCA
G I V Q S L T Y A L S V S F V N L N F A
20941 TCGCCTACCTTTACGGTATCTGGCTGGTTGGGAGGAGAATCTGCTCTCCGTATACGGTGT
I A Y L Y G I W L V G R R I C S P Y T V
21001 TCCA_
F Q
21005 gtaagttcaaagcatttccatactcttgccggtgagaaagagtctcattagttctt
21061 cag
21064 __GGTGATTGAATCCTTGAACACCGCGTCGATGTCTCTGATTGCATTTGCCACTTACTT
V I E S L N T A S M S L I A F A T Y F
21121 CCCTGAGTATGTCCGTGCTCGACTATCAGCAGCCCTTCTTTTCCGAATGCTTAGGGACAA
P E Y V R A R L S A A L L F R M L R D K
21181 ACCGAAAATCGACAGCCTGTCCCCATTAGGAATGCAAACT
P K I D S L S P L G M Q T
21221 gtgagttttgaccaaatttt
21241 ttggcgagccttgtagcatattatcgatttttgcag
21277 AAGTTACGCGGATCAATTCACTTT
K L R G S I H F
21301 TCTGATCTGTCCTTCTCATATCCCGTCAGTCGTAGAGATATGGTACTCAAAGGAATCACT
S D L S F S Y P V S R R D M V L K G I T
21361 CTTAAGGTAATTACTGGCGAATCCA__
L K V I T G E S X
21386 gtacgaaaatgctatttctggaaaactaaaccaca
21421 aaaatgttttacaaagttctgtcggacttaatttctaaaccaagaaggaattcttccaac
21481 gaaattatcttgaaatttcctcctcatcctagacaaataataggaaatttatcttttgcg
21541 aatggagagaaagagcgatcaggacctgctacagagcaaacctctgacaattaaagaaat
21601 taacacaaaattgagatagtcttgtgggagtggggacgaatttcgtgggcgctagaggat
21661 agacgctcagcctgaagactcaaaactctttctagaattgaaacttcatctttgaaaagc
21721 agacccaagattagctatgagaaacgggtttgcggtctttgaaggcgtcactttgagctt
21781 gagttttgttaactcttctctgaactctttctaggtccctgcagggaagtactacgaaaa
21841 ctaccggtcccgctttcatttgtttgtgctcnnnnnnnnnnnnnnnnnnnnnnnnnnnn
21901 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
21961 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
22021 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
22081 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
22141 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
22201 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
22261 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
22321 nnnnnnnnnnnnnnnnnnnnnnnnnnnnn
X X X X X X X X X X X X X X X X X X X X



```
27301    atacggtaggacctaaagcaacgtatcaagctaaattagagaagatattccctggaatat
27361    ctattacggtaatttttcaagtttttacttctcttccttctcgtttggtcctgtagtgaa
27421    ataatggcatacattttggacgtgagaaatgctgtttagcctgctggctgagctgatatt
27481    catagccctttgtgctcgctatggtcgcgcattccaagccatcggctgctatgcttgtg
27541    atttttcccttgtcccagtcattgtcatcgagaaactttcaagagagcccaatttttta
27601    ggtaacggagaaggctgacgccctcttcccaattgttagcgctgcttctatagctgctaa
27661    agtgacacgtgatcgtcgccttcggtcatggaaattcgttgagcccgatgtaaaaattcc
27721    agctgacggttacggtagtggttacccccggaggttagctttttttcagagctcgattagtc
27781    ttgatcttgcagtatttcttggatcatgcttctttatattatccacctgttttcagatcc
27841    taataccaagaagtttctggtggatagtgtggatccaatatttggatattcatctttagt
27901    acgattttcatggaaaactgctgaagttttagttgataagaattgtgtgaaagcagaatg
27961    gcaagttgaaaacattgaaagatttaatttctagttttaagtcaaaaattttagggaaga
28021    acaagaagccaaggcgccttcagtgaaaggttggcttactacaaaactggagttgcctaa
28081    aaggcatgtgtattatgcagataggactattcagaatgttgcgtctttctaaacgtgaac
28141    ggctcagtactaatttcctttctttttttttatttggttatttcactcgaatgatgttgt
28201    gatctccacagcttttctttgttttatacgcctttctcatctccgtatgtatcaattttt
28261    gaaggttgctcatctagtggtactattgaaataaaggctttatttttcgaatcgctgttt
28321    ccatcttgacggtattattccgtagtatctcactatttccccttaaaagtgctaactagc
28381    tctatctttgagcatagtttcagatggggtcttggttcagcctatcgacgattttaactc
28441    ttttgactgcatattatattggtcatatgataaatactttctataaaatttattcggtac
28501    cattttgtaaggaaggtcttgcgtgtgggaaggtaggttttttaaaccttttataaaagca
28561    ggttcatactacaaaacatatatgcagacaaattcatcgttctttcgttacgtttttcag
28621    cctcttctacctcttcaaaagaatgggaaatttccgttggttcagttccgtgtatatgct
28681    tctgataaaagtttctccatagggaaagttggagaaaatgataaacaattttcagctggct
28741    gatccacaagacatcgaggtacgannnnnnnnnnggacaaaccgaaaatcgacagcctgt
28801    ccccattaggaatgcaaactgtgagttttgaccaaattntttggcgagccttgtagcata
28861    ttatcgatttttgcagaagttacgcggatcaattcacttttctgatctgtccttctcata
28921    tcccgtcagtcgtagagatatggtactcaaaggaatcactcttaaggtaattactggcga
28981    atccagtacgaaaatgctatttctggaaaactaaaccacaaaaatgttttacaaagttct
29041    gtcggacttaatttctaaaccaagaaggaattcttccaacgaaattatcttgaaatttcc
29101    tcctcatcctagacaaataatggaaatttatcttttgcgaatggagagaaagagcgatc
29161    aggacctgctacagagcaaacctctgacaattaaagaaattaacacaaaattgagatagt
29221    cttgtgggagtggggacgaatttcgtgggcgctagagggtagacgctcagcctgaagact
29281    caaaactctttctagaattgaaacttcatctttgaaaagcagacccaagattagctatga
29341    gaaacgggtttgcggtctttgaaggcgtcactttgagcttgagttttgttaactcttctc
29401    tgaactctttctaggtccctgcagggaagacggttgcattggttggacctagcggttgtg
29461    gaaaaagcacttctattcaacttatcgaacgattctacgatcctgtggctggatctgtgg
29521    tgaggctttttgagggagatgacatccaccaaacggcttcccatagattaacagtactt
29581    tcggtatgcaggatgggtttacctaaccttataaaggtgcacttaaaaacgttttccttt
29641    ctcctgtccacaaatcgcattcgactcgtccatttctttcatgaaatgcatttcttactg
29701    tagataaagctgggtgctgcggatccaaacgaataatctgagctttgcagcttttgatga
29761    agtagacgctcgtgaactgaatttgaggcatctccgttcacaaatgtcacttgtcggaca
29821    agaacctattctttcaactattccatcaggtacgaacttttcttacctttcctaactaaa
29881    cagaaaacttactgaacagaaaactttgtagagaaaacattgcgtacggcctcgagcaa
29941    gcgactgttgatcagattgaaaatgcagctaagttggcgaatgcccacaacttcatcatc
30001    a
```

Sequence 1: Annotated *H. contortus supercontig* 0001880 from S09 Database.

Full nucleotide sequence of *supercontig* aligned to *Cel*-PGP-4 with translation of putative exons. Green highlight: BLAST and manually edited alignments of *H. contortus* DNA to *C. elegans* protein. Yellow highlight: unknown sequence between *contigs* within *supercontig*. Red highlight: *C. elegans* amino acid sequence unmatched.

All *supercontigs* implicated in the preceding stages were investigated using the BLAST method but only one example is given above. The sections of unknown sequence and regions of hypervariability were handicaps to determining full-length nucleotide and amino acid sequence as shown by the regions of red and yellow highlights so an improvement was sought using gene structure prediction algorithms as shown below in Sequence 2.

3.3.1 Gene Structure Prediction Algorithms

The successful gene structure prediction algorithm in this study was the GeneWise algorithm. The output for the same *supercontig* as in

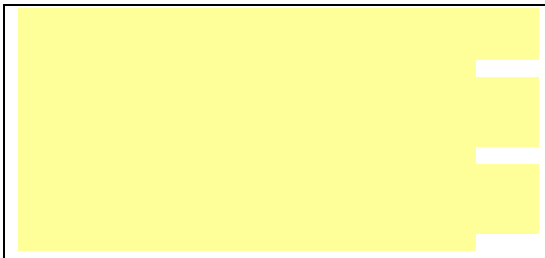
Sequence 1 is shown in Sequence 2 below. The entire set of outputs for the 48 *supercontigs* from the S09 database with evidence of *pgp* exons are in the Appendix.

```
gene wise
Query protein:       CAEEL-PGP-4_1a_CAEEL-PGP-4_F42E11_1a_1-1280AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            global
Target Sequence      H_S09_0001880___HAECO_S09_Supercontig_0001880
                                    _1-30001bp
Strand:              both
Start/End (protein)  global
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623

H_S09_0001880__     1          Intron     CAG
                    <0-----[     :13980]-0>
                     449 n [5271 : 5719]
                     234 n [7717 : 7950]
                      10 n [9514 : 9523]
                     963 n [10866:11828]
                      10 n [12734:12743]


CAEEL-PGP-4_1a_    47 FRHSGCADYLLLLGGLVLSAANGALLPFNSLIFEGITNVLMK
                      FR++   D+ L+L G V +A  G     +SL+F  + + L+
                      FRYATKLDFCLMLLGAVFAATQGTFNSVSSLVFRHLMDALII
H_S09_0001880__13981 tctgaacgtttatcgggtggacgatatgttcgtcccaggcaa
                      tgaccatatgttttgcttcccagctactcctttgattacttt
                      tctaaattctggagaggtactaatcttaatgttgttgttaca


CAEEL-PGP-4_1a_    89              G          EAQWQNGTFDYDTFSSGIQHYCL
                                   G          E +WQ G FD   F+    +
                                   G          EFEWQAGIFDDYEFTQLAMNSVY
H_S09_0001880__14107 g          [ggt]        gtgtcggatggtgtaccgaatgt
                      GTGGAAC  Intron 1   TAGgatagacgttaaaatcatctacta
                      <1-----[14108:14173]-1>tacagatcaccttgcgagtgttct


CAEEL-PGP-4_1a_   113 LYFLLGVLMFTCTYFS                       NACLFTMAER
                       Y L G+++FT  ++S                        C  T+ ER
                      RYTLFGLIQFTLGFLS                       MCCWHTVCER
H_S09_0001880__14245 atactgcactacgtct                       atttcagtgc
                      gacttgttatctgttcGTCAGTT  Intron 2   TAGtgggactgag
                      actattcaacttatag<0-----[14293:14794]-0>gttgtaataa


CAEEL-PGP-4_1a_   139 RLYCIRKHLLQSVLRQDAKWFDENTVGGLTQKMS
                      ++Y IR     SV+RQD  WFD+N  G LT +MS
                      QVYQIRNRFFGSVIRQDMAWFDQNDSGALTTRMS
H_S09_0001880__14825 cgtcaaacttgtgaccgagttgcagaggcaacaa
                      ataatgagttgcttgaatcgtaaaaggctccgtg
                      aatacatcttaagaagtgcgtcattcatgatggc


CAEEL-PGP-4_1a_   173              S          GIEKIKDGIGDKIGVLVSGIATF
                                   GI++I+DGIGDK+  + + +ATF
                                   D          GIDRIRDGIGDKLDAMFAYFATF
H_S09_0001880__14927 g          [gat]        gagcacggaggacggatgttgat
                      aGTATGTT  Intron 3   CAG gtagtgagtgaatacttcatcct
                      <2-----[14929:15689]-2>taatgagttcccacttgcgttcgc


CAEEL-PGP-4_1a_   197 ISGVALGFYM              C          WQLTLVMLVTVPL
                      I+G+ + +                           WQ+TLVM+   P+
                      IAGITVALSC              S          WQMTLVMIGFFPI
H_S09_0001880__15760 aggaaggcata           [agt]        tcaacgaagttca
                      tcgtctctgggGTAAGAA  Intron 4   CAG gatcttttgttct
                      ctaaattgtt <2-----[15792:15905]-2>tgagtgtgtgtcac


CAEEL-PGP-4_1a_   221 QLGSMYLSAK                       HLNRATKNEMSAYSSA
                        G + +++                        + +    E+  Y  A
                      FFGPLTVTSM                       IMGKVVPKEQEFYVRA
H_S09_0001880__15946 ttgccagata                       aagaggcagcgttgcg
                      ttgctctcctGTGAGGG  Intron 5   CAGttgattcaaaatatgc
                      ctcactatag<0-----[15976:16029]-0>tgagcacagagtctgg


CAEEL-PGP-4_1a_   247 GGMANEVIAGIRTVIAFNAQPFEI                      270
                      G  A EV+ GIRTV+AFN Q  EI
                      GSTAEEVVNGIRTVVAFNGQEKEI
H_S09_0001880__16078 gtagggggagacagggtagcgaga                    16149
                      gcccaattagtgcttctagaaaat
                      ctctagacccactgacttaggagt
                                  884 n [16150:17033]


CAEEL-PGP-4_1a_   271 ERYGAQLAKARKMGIRKAIVLALCSAMPLFLMFVLMAGAFWYGAILTSYG
H_S09_0001880__       ----------------------------------------------------


CAEEL-PGP-4_1a_   321 VATSGTTFGVFWAVILGTRRLGEAAPHMGAITGARLAVNDIFKVIDHEPE
H_S09_0001880__       ----------------------------------------------------


CAEEL-PGP-4_1a_   371 INCTKQEGRRPDKVNGKLVFDNIQFTYPTRPDVKILKGVSFEVNPGETIA
```

```
H_S09_0001880__       ------------------------------------------------

CAEEL-PGP-4_1a_  421 LVGHSGCGKSTSIGLLMRFYNQCAGSIKLDGIPIEDYNIQWLRSTIGIVQ
H_S09_0001880__       ------------------------------------------------

CAEEL-PGP-4_1a_  471 QEPIIFLATVAENVRMGDDSITDKDIENACRQANAHDFIGKLSE   514
H_S09_0001880__       ---------------------------------------------

H_S09_0001880__              Intron     CAG                      17518
                      <0-----[     :17518]-0>

CAEEL-PGP-4_1a_  515 GYNTVIGAGAVQLSGGQKQRVAIARAIVRKPQILLLDEATSALDTE
                      GY T IG G VQLSGGQKQRVAIARA+VR P+ILLLDEATSALD E
                      GYQTKIGEGGVQLSGGQKQRVAIARALVRNPRILLLDEATSALDAE
H_S09_0001880__17519 gtcaaaggggcttggcaccggagcgcgaaccacccgggaagtggg
                      gaacatgaggtatcggaaagtctcgcttgacgttttaaccgctaca
                      ccgaaaaatatgactaaggattctacgtgtaaatagtgagttgtaa

CAEEL-PGP-4_1a_  561 SERMVQTALDK                        ASEGRTTLCIAHRLS
                      SE +VQ AL+                         A  GRTT+ IAHRLS
                      SESIVQQALEN                        AQSGRTTISIAHRLS
H_S09_0001880__17657 agaagccgcga                         gctgaaaatagcact
                      gagttaactaaGTACGTT  Intron 7   CAGcacggcctctcagtc
                      cattcaggtat<0-----[17690:18028]-0>tacggcgtgcatatg

CAEEL-PGP-4_1a_  587 TIRNASKILVFDQGLIPER              G          IHDQ
                      TI+N  +I VF+ G I E               G           HD+
                      TIKNVDRIYVFNNGRIVED              G          NHDE
H_S09_0001880__18074 aaaaggcatgtaagaagggg           [gga]        acgg
                      ctaatagtattaaggttaa GTGAGCT  Intron 8   TAGgaaaa
                      aaatctcccaccttgagac <1-----[18132:18212]-1>accca

CAEEL-PGP-4_1a_  611 LIRQNGIYANMVRAQEIEKAKDDTTQDDDELV
                      L++ NG+Y+ +VRAQEIE+     +  DDE
                      LMKMNGLYSELVRAQEIEQL--EKSGSDDE-T
H_S09_0001880__18227 caaaagtttgtgagcgagcc  gatgaggg a
                      ttatagtacattgcaataat  aacggaaa c
                      cgagcggcggggagagtgat  gatgccta t

CAEEL-PGP-4_1a_  643              E          EDNYSISRRLSTSEEELRKSKSL
                                            E N ++ RR          SK L
                                 A          EHNVTLMRR---------RSKRL
H_S09_0001880__18314 g          [gct]       gcagacaaa         aaacc
                       GTGAGAG  Intron 9   CAGcaaatcttgg         ggagt
                       <1-----[18315:18383]-1>tactgaggga         acgcc

CAEEL-PGP-4_1a_  667 LRDSTRFSQSMLSVTSQVPDWEMESAR
                       R  +R         +++   E+E+
                      SRSISR--------PTELRGQELENLE
H_S09_0001880__18428 tctatc        cagccgcgtgacg
                      cgctcg        ccatggaataataGTAACCT  Intron 10
                      ctccat        cgatttagaaccg<0-----[18485:18570]

CAEEL-PGP-4_1a_  694    EEMIEEGAMEASMMDIFRFAKPEKMNIVIALIFTLIRGITWPAFSV
                        EE+ E+    AS++DI++FA+ E   + +ALI+ + RG+T+P FS+
                        EEVEEKKVKGASLLDILKFARQEWCQLAVALILAVARGMTFPVFSI
H_S09_0001880__18568    gggggaagaggaccgacatgacgttccgggcacgggcgaatcgtta
                      TAGaataaaatagcgttattatcgaaggatctctttctcggtctcttct
                      -0>aacggaagaaattattggtaagagcagagctcttattgggctgtga

CAEEL-PGP-4_1a_  740 VYGQLFKVFAEGGEDLPVNA              -          ---
                      +YGQ+FKV A      P+N
                      IYGQMFKVRAFRRLLFPMNY              G          NSQ
H_S09_0001880__18709 atgcatagagtactctcaatg           [ggc]        atc
                      tagattatgctggtttctaagGTTAATT  Intron 11  CAG aca
                      tttagcggaacgtggctgct <2-----[18771:19450]-2>ccca

CAEEL-PGP-4_1a_  760 -----LISSLWFVLLAVTSAVTTFISGSLLGKTGETMSSRLRMDVFKNI
                           ++++WF LL ++S ++T+ISG L G+ GE++++RLR+ +F NI
                      KLHGATMNAIWFSLLGISSGISTMISGYLFGRIGESLTNRLRLSLFTNI
H_S09_0001880__19461 accggaaagatttccgatagaaaaatgtctgaaggtcaaaccctctaaa
                      atagcctactgtcttgtcggtgcttcgattggtgactcagtgtcttcat
                      gatacggtgagcaagctttacagttctgtgacaataataatacgtgtc

CAEEL-PGP-4_1a_  804 MQQDAT              -          YFDDPKHNVGNLTSRLA
                      ++Q +                         YFD   H  G LT+RLA
                      VKQVSK             E          YFDHEDHASGKLTTRLA
H_S09_0001880__19608 gacgaag           [gag]        ttgcggcgtgataaatg
                      taatga GTCAGAA  Intron 12  GAGaataaaaaccgatccgtc
                      agaacg <1-----[19627:19682]-1>gctccatttaaagcaagg

CAEEL-PGP-4_1a_  827 TDSQNVQA                        AIDHRLAEVLNGVVSLFT
                      TD+ N++A                        AID RLA+V++ V S++
                      TDAPNIRA                        AIDQRLADVVSAVSSIIG
H_S09_0001880__19736 aggcaaag                        gagcccggggtggttaag
                      caccatgcGTATGGT  Intron 13  TAGctaagtcattcctccttg
                      gcaattag<0-----[19760:20114]-0>actattttccatgggata
                               10 n [20086:20095]

CAEEL-PGP-4_1a_  853 GIAVAFWFGWSMAPIGLITAL                       LLVIA
                      GI++AF +G +MAPIG++TA+                        L+
                      GISIAFSYGPAMAPIGVLTAV                       TLITL
H_S09_0001880__20169 gatagtttgcgagcaggcagg                        acaac
                      gtctctcagcctcctgttcctGTAAGAA  Intron 14  CAGcttct
                      ctcaatccataggatactgca<0-----[20232:20289]-0>tacag

CAEEL-PGP-4_1a_  879 QSAVAQYLKYRGPKDMESAIEASR                      IV
```


Q+ VA+YLK RG +D A E SR +
QTLVARYLKVRGQRDAVLAEEPSR LA
H_S09_0001880__20305 cacggctcagcgccgggcgggcta tg
acttcgatatggagacttcaaccgGTAAGCC Intron 15 CAGtc
aattcatggcaaattagaagagaa<0-----[20377:20442]-0>gt
CAEEL-PGP-4_1a_ 905 TESISNWKTVQALTKQEYMFHAFTAASKNPRKRAFTK
TE+I KTVQ LTK+ ++ F P KRA+ +
TEAIEQHKTVQYLTKERQFLDKFVTQMHGPHKRAIFR
H_S09_0001880__20449 aggagccaagctcaagccttgatgacacgccaagatc
cactaaaactaatcaagattaattcatagcaagcttg
gagaagtgacacatgaggcgtactcagtttcagaata
CAEEL-PGP-4_1a_ 942 G LWQSLSFALAGSFFLWNFAIAYM
G + QSL++AL+ SF NFAIAY+
G IVQSLTYALSVSFVNLNFAIAYL
H_S09_0001880__20560 g [ggt] agcttatgctgatgatatgagtc
GTACTGA Intron 16 CAGgttactcactctgttatatctcat
<1-----[20561:20880]-1>tttgagttccctttccgctcccct
CAEEL-PGP-4_1a_ 966 FGLWLISNNWTTPFAVF Q VIEALN
+G+WL+ +P+ VF Q VIE+LN
YGIWLVGRRICSPYTVF Q VIESLN
H_S09_0001880__20952 tgatcggaaattctagtc [cag] gagtta
agtgttgggtgccacttaGTAAGTT Intron 17 CAG ttacta
ctcggtggacctgtggc <2-----[21005:21063]-2>ggtacgc
CAEEL-PGP-4_1a_ 990 MASMSVMMAASYFPEYVRARISAGIMFTMIRQKAKIDNRGLTGETP
ASMS++ A+YFPEYVRAR+SA ++F M+R K KID+ G
TASMSLIAFATYFPEYVRARLSAALLFRMLRDKPKIDSLSPLGMQT
H_S09_0001880__21083 agtatcagtgattcgtgcgcctggcctcacagacaagactctgaca
ccctcttctccatcaatgcgtccctttgttgaacatagtcctgtac
cgggtgtatctcctgtcttaaaacttcagtgcagacccgcaaagat
CAEEL-PGP-4_1a_ 1036 DIRGDISMKGVYFAYPNRN
+RG I + + F+YP
KLRGSIHFSDLSFSYPVSR
H_S09_0001880__21221 atcgtacttgcttttcgac
GTGAGTT Intron 18 CAGatggctatcatctcactgg
<0-----[21221:21276]-0>gacaatctttgccatcctt
CAEEL-PGP-4_1a_ 1055 RQLILNNFNMSAQFGET 1071
R ++L + + GE+
RDMVLKGITLKVITGES
H_S09_0001880__21334 agagcagaacagaaggta 21385
gatttagtctattcgac
atgacaacttgattcac
H_S09_0001880__21386 GTACGAA Intron
<1-----[21386:]-1>
480 n [21872:22351]
CAEEL-PGP-4_1a_ 1072 VALVGPSGCGKSTSIQLIERYYDAICGAVKIDDHDIRDISVKHLRHNIAL
H_S09_0001880__ ----------
CAEEL-PGP-4_1a_ 1122 VGQEPTLFNLTIRENITYGLENVSQEQVEKAATLANIHSFVENLPE1167
H_S09_0001880__ ----------
CAEEL-PGP-4_1a_ 1168 GYDTSVGASGGRLSGG
GYDT VG G LSGG
GYDTIVGERGSMLSGG
H_S09_0001880__22353 gtgaagggcgaattgg
gaacttgagggttcgg
ctcgtttgactggaca
CAEEL-PGP-4_1a_ 1184 QKQRIAIARAIVRNPKILLLDEATSALDTESEK
QKQRIAIARA++R+PKILLLDEATSALDTESEK
QKQRIAIARAVIRDPKILLLDEATSALDTESEK
H_S09_0001880__22401 caccagagcggacgcaatctgggaagtgagaga
aaagtctcgcttgacattttaaccgctacagaa
gagaccaaattcttaatacatatgtgattataa
CAEEL-PGP-4_1a_ 1217 IVQEALDKARLGRTCVVIAHRLSTIQ
IVQEAL+KAR GRTC+VIAHRLS+IQ
IVQEALEKARQGRTCIVIAHRLSSIQ
H_S09_0001880__22500 agcggcgagacgcatagagcccttac
GTACGTC Intron 20 CAGttaactaacgaggcgtttcagtccta
<0-----[22500:22554]-0>taaaggaagaatatttccttagttta
CAEEL-PGP-4_1a_ 1243 NADKIIVCRNGKAI EEGTHQTLLARR
NAD IIV ++G E+GTHQ LLAR
NADLIIVIKDGMVE EQGTHQQLLARE
H_S09_0001880__22633 aggcaagaaggagg gcgaccctcgag
acatttttaagttaGTAAGCT Intron 21 CAGaagcaaattcga
tattccctacagtg<0-----[22675:23148]-0>gagacgaactaa
CAEEL-PGP-4_1a_ 1269 GLYYRLVEKQST- 1280
GLY +V KQ
GLYASMVTKQDLK
H_S09_0001880__23185 gctgaagaacgca 23223
gtacgttcaaata
catacgacaatcg
CAEEL-PGP-4_1a_ *
*
*
H_S09_0001880__23224 t 23226
a

```
                    a


H_S09_0001880__23227       [23227:29424]                          29424
                     10 n [23649:23658]
                     10 n [28765:28774]
                      1 n [28839:28839]
//

Gene 1
EXONS 13981 23223
  Exon 13981 14107 phase 0
  Exon 14174 14292 phase 1
  Exon 14795 14928 phase 0
  Exon 15690 15791 phase 2
  Exon 15906 15975 phase 2
  Exon 16030 16149 phase 0

  ---

  Exon 17519 17689 phase 0
  Exon 18029 18131 phase 0
  Exon 18213 18314 phase 1
  Exon 18384 18484 phase 1
  Exon 18571 18770 phase 0
  Exon 19451 19626 phase 2
  Exon 19683 19759 phase 1
  Exon 20115 20231 phase 0
  Exon 20290 20376 phase 0
  Exon 20443 20560 phase 0
  Exon 20881 21004 phase 1
  Exon 21064 21220 phase 2
  Exon 21277 21385 phase 0

  ---

  Exon 22353 22499 phase 2
  Exon 22555 22674 phase 0
  Exon 23149 23223 phase 0
//

Making a G in phase 1 intron
Making a D in phase 2 intron
Making a S in phase 2 intron
Making a G in phase 1 intron
Making a A in phase 1 intron
Making a G in phase 2 intron
Making a E in phase 1 intron
Making a G in phase 1 intron
Making a Q in phase 2 intron


>H_S09_0001880___HAECO_S09_Supercontig_0001880
_13981-16149bp_AA
---
_17519-21385bp_AA
---
_22353-23223bp_AA
_CAEEL-PGP-3_OR_4_CAEEL-PGP-4_1a_47-270AA---515-1071AA---1168-1280AA

FRYATKLDFCLMLLGAVFA
ATQGTFNSVSSLVFRHLMDALIIGEFEWQAGIFDDYEFTQLAMNSVYRYTLFGLIQFTLG
FLSMCCWHTVCERQVYQIRNRFFGSVIRQDMAWFDQNDSGALTTRMSDGIDRIRDGIGDK
LDAMFAYFATFIAGITVALSCSWQMTLVMIGFFPIFFGPLTVTSMIMGKVVPKEQEFYVR
AGSTAEEVVNGIRTVVAFNGQEKEI

GYQTKI
GEGGVQLSGGQKQRVAIARALVRNPRILLLDEATSALDAESESIVQQALENAQSGRTTIS
IAHRLSTIKNVDRIYVFNNGRIVEDGNHDELMKMNGLYSELVRAQEIEQLEKSGSDDETA
EHNVTLMRRRSKRLSRSISRPTELRGQELENLEEEVEEKKVKGASLLDILKFARQEWCQL
AVALILAVARGMTFPVFSIIYGQMFKVRAFRRLLFPMNYGNSQKLHGATMNAIWFSLLGI
SSGISTMISGYLFGRIGESLTNRLRLSLFTNIVKQVSKEYFDHEDHASGKLTTRLATDAP
NIRAAIDQRLADVVSAVSSIIGGISIAFSYGPAMAPIGVLTAVTLITLQTLVARYLKVRG
QRDAVLAEEPSRLATEAIEQHKTVQYLTKERQFLDKFVTQMHGPHKRAIFRGIVQSLTYA
LSVSFVNLNFAIAYLYGIWLVGRRICSPYTVFQVIESLNTASMSLIAFATYFPEYVRARL
SAALLFRMLRDKPKIDSLSPLGMQTKLRGSIHFSDLSFSYPVSRRDMVLKG
ITLKVITGES

GYDTIVGERGSMLSGGQKQRIAIARAVIRDPKILLLDEATS
ALDTESEKIVQEALEKARQGRTCIVIAHRLSSIQNADLIIVIKDGMVEEQGTHQQLLARE
GLYASMVTKQDLK
//

>H_S09_0001880___HAECO_S09_Supercontig_0001880
_13981-16149bp
---
_17519-21385bp
---
_22353-23223bp

tttcgctatgcaacaaaacttgatttctgtttgatgttactgggagcggtgtttgcagcc
actcaaggaactttcaattctgtatcatctctggtttttcggcatcttatggatgctcta
atcataggtgaattcgaatggcaagctggcatattcgacgattatgagttcacgcaactg
gctatgaattctgtctatagatacactctatttggtctcatacaattcactcttggattt
ctatcgatgtgttgttggcatacagtatgtgaacgacaagtatatcaaatcagaaatcgc
ttttttggatcagtgatacgacaggatatggcctggtttgaccaaaatgatagcggagct
ctgacaactcggatgagcgatggaatagatcggatacgggatggtatcggcgacaaactc
gatgctatgttcgcgtattttgccacgttcatcgctggaataacagttgctctgagttgt
agttggcaaatgactctggttatgattgggtttttcccaatcttctttggcccactcact
gtaacttcaatgattatgggaaaggtcgtacccaaagagcaagagttttacgttcgggcg
ggctctaccgctgaagaggtagtcaacggcatacgcactgtggtagcctttaatggacag
gagaaagagatt
```



```
ggctaccagacaaaaataggagaaggtggagttcagttatccggtggacaaaagcag
cgagttgctatcgctcgagccctggttaggaatccacgaatacttctactggatgaggca
acgagtgctttggatgcagaaagcgaaagtattgtccaacaggcgcttgaaaatgctcaa
tccgggaggaccacgatttcgatcgcacatagactttcgacaataaaaaatgtcgatcgc
atctacgtattcaacaatggtaggatagtggaagacggaaaccacgacgaactcatgaaa
atgaacgggttgtactcggagttggtgagagcgcaagagattgagcaacttgagaaatct
gggagcgacgatgaaactgctgaacacaatgtgacactgatgaggagaagaagcaagcgc
ctctcccgttccatctcacgtcccacggaacttcgtggtcaagagttagaaaacctcgag
gaagaagtcgaggagaaaaaagtgaaaggagcaagtcttctagatattctgaagtttgca
agacaggaatggtgccaactggcagtggcccttatccttgctgtagctcgtgggatgacg
ttccctgtgttttcgataatttatggtcaaatgttcaaggtgagagcattcaggcgtttg
ctgttccctatgaactatggcaactcccaaaagctacatggagccacgatgaatgcgata
tggttctcactactgggcatttctagtggtataagcacaatgatttctggctatctgttt
gggagaatcggagaatctctaacaaatagactacgtctatccctgtttacgaatatcgta
aagcaagtaagcaaggagtactttgaccacgaagatcatgcttcaggaaaattgaccaca
agattggcgacggacgcaccaaatattagagcggcaatcgatcaacgtcttgctgatgtc
gtctcagctgtgtcgtcgataattggaggcatttccatagcattttcctacggacctgca
atggcgccaattggagtccttacggccgtaactctaatcacactgcaaacacttgttgcc
cgatatctgaaggtccgaggacaacgtgatgcagtgctagcagaggaaccgtcaagattg
gctacggaagcgatagaacagcataagacagtccaatacctaactaaggaacggcagttc
ttggataaattcgttacccaaatgcatggtcctcacaaaagggcaatatttcgaggtatt
gttcagtcattgacttatgccctctccgttagttttgtcaacttgaactttgccatcgcc
taccttacggtatctggctggttgggaggagaatctgctctccgtatacggtgttccag
gtgattgaatccttgaacaccgcgtcgatgtctctgattgcatttgccacttacttccct
gagtatgtccgtgctcgactatcagcagcccttctttccgaatgcttagggacaaaccg
aaaatcgacagcctgtccccattaggaatgcaaactaagttacgcggatcaattcacttt
tctgatctgtccttctcatatcccgtcagtcgtagagatatggtactcaaagga
atcactcttaaggtaattactggcgaatcca
```



```
ggctatgacacgattgttggtgagcgaggcagtatgttgtca
ggcggacagaaacagcgaatcgccatagcacgagctgttatccgtgatccaaaaatttta
ctcttagatgaagctacgagtgcgttagatactgaaagtgaaaaaattgtacaagaagcg
ctggaaaaagcgagacaaggtcgaacttgtattgtcatcgctcatcgactgtcttctatt
caaaatgcagatcttatcatcgtcattaaagacggaatggttgaggagcaagggacacac
cagcaattactcgctagagaagggcctatatgcaagcatggtaaccaaacaagatctcaag
//
genewise
Query protein:       CAEEL-PGP-4_1a_CAEEL-PGP-4_F42E11_1a_1-1280AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0001880___HAECO_S09_Supercontig_0001880
                                    _23227-30001bp
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623
H_S09_0001880__ 23227          Intron     CAG
                     <2-----[     :29424]-2>
                       10 n [23649:23658]
                       10 n [28765:28774]
                        1 n [28839:28839]


CAEEL-PGP-4_1a_ 1070  ETVALVGPSGCGKSTSIQLIERYYDAICGA               1099
                      +TVALVGPSGCGKSTSIQLIER+YD + G+
                      KTVALVGPSGCGKSTSIQLIERFYDPVAGS
H_S09_0001880__29425  aaggtggcagtgaaataccagcttgcgggt               29516
                      gactcttgcggggagcctattagtaactcgc
                      gggtagtatcttaactttatcaaccttgtat


CAEEL-PGP-4_1a_ 1100 V                                            1100

                     !!
H_S09_0001880__29517 gt                                          29518


CAEEL-PGP-4_1a_ 1101                K          IDDHDIRDISVKHLRHNIALVGQ
                                               +D+ D R+++++HLR  ++LVGQ
                                    A          FDEVDARELNLRHLRSQMSLVGQ
H_S09_0001880__29519 g           [gct]         tgggggcgcataccctcatcggc
                      GTGAGGC  Intron 1   CAGctaatacgatatgatgcatcttga
                      <1-----[29520:29750]-1>tttaacttagtggtctaagatcaa


CAEEL-PGP-4_1a_ 1125 EPTLFNLTI              -          ---RENITYGLENV
                     EP L  +                              RENI YGLE
                     EPILSTIPS              E          NFCRENIAYGLEQA
H_S09_0001880__29822 gcactaactg           [gaa]         attagaagtgcgcg
                     acttcctcc GTACGAA  Intron 2   CAGaatggaatcagtaac
                     atttattaa <1-----[29850:29899]-1>acttaactgcccgag
```

```
CAEEL-PGP-4_1a_ 1145 SQEQVEKAATLANIHSFVE                          1163
                     + +Q+E AA LAN H+F+
                     TVDQIENAAKLANAHNFII
H_S09_0001880__29944 aggcagaggatgagcataaa                       30001
                     ctaataaccatcacaattt
                     tttgtatatgggtccccc
//

Gene 1
EXONS 29425 30001
  Exon 29425 29516
       29517 29518
       29519 29519 phase 0
  Exon 29751 29849 phase 0
  Exon 29900 30001 phase 1
//

Making a A in phase 1 intron
Making a E in phase 1 intron


>H_S09_0001880___HAECO_S09_Supercontig_0001880
_29425-29516bp_AA
_29517-29518bp_Frameshift
_29519-30001bp_AA
_CAEEL-PGP-3_OR_4_CAEEL-PGP-4_1a_1070-1099AA_1100AA_Frameshift_1101-1163AA

KTVALVGPSGCGKSTSIQLIERFYDPVAGS


AFDEVDARELNLRHLRSQMSLVGQEPIL
STIPSENFCRENIAYGLEQATVDQIENAAKLANAHNFII
//

>H_S09_0001880___HAECO_S09_Supercontig_0001880
_29425-29516bp
_29517-29518bp_Frameshift
_29519-30001bp

ggaagacggttgcattggttggacctagcggttgtggaaaaagcacttctattcaactt
atcgaacgattctacgatcctgtggctggatct

gt

gct
tttgatgaagtagacgctcgtgaactgaatttgaggcatctccgttcacaaatgtcactt
gtcggacaagaacctattctttcaactattccatcagaaaacttttgtagagaaaacatt
gcgtacggcctcgagcaagcgactgttgatcagattgaaaatgcagctaagttggcgaat
gcccacaacttcatcatca
//
```

Sequence 2: Annotated *H. contortus supercontig* 0001880 from S09 Database using the GeneWise Algorithm.

GeneWise parameters: previously established as homologous protein for the putative transcript from *H. contortus supercontig* 0001880. Global model parameter in GeneWise generated greatest alignment. Alignment starts an estimated 47 codons from start; residues form positions 271-514 and 1072-1167 not matched; alignment extends to stop signal. Alignment of *C. elegans* residues, translated codons and nucleotides provided then list of base positions of exons, list of introns in phase 1 or 2, split codons, translation of putative *H. contortus* sequence, transcript of putative *H. contortus* sequence. Additional alignment to *C. elegans* residue positions 1070-1163 provided, despite including a frame shift. Green highlight: alignment. Yellow highlight: unknown sequence between *contigs* within *supercontig*.

GeneWise picked up two additional exons (starting at nucleotide position 13981 in Sequence 2) 5' to the first exon (starting at nucleotide position 14795 in Sequence 1) detected by the BLAST method. However, the start of the gene was not detected by either method. GeneWise was also sensitive to the exon starting at position 15690, missed by BLAST. Where no identities existed for a stretch of 884 and then 480 nucleotides no method could be effective. GeneWise predicted an exon (starting at position 18384) with weak homology and two strings of 8 and 9 residues in the *C. elegans* sequence with no alignment to *H contortus* whilst the BLAST method just indicates a lack of alignment. This is in the region between

domains that does commonly demonstrate variable length between proteins and species.

The number of sequences from the S09 database that were aligned with E-values below 10, was between 57 (against *Cel*-PGP-11) and 73 (against *Cel*-PGP-2). The sequence on the *forward BLAST* that was lowest in the order of E-values that still provided evidence of *pgp* coverage on the *back BLAST* was the 47th sequence out of 67 sequences (against *Cel*-PGP-10). This provided evidence that an E-value of 10 as a cut-off point was still sufficiently sensitive. The total number of sequences that were aligned to *pgp* genes with an E-value below 10 during the *forward BLAST* was 95. The number with better alignment to *pgp* rather than other gene families (over part or all of the sequence) on the *back BLAST* was 48. The remaining 47 sequences were considered by this evidence to be false positives from the *forward BLAST*.

Manually editing the alignments generated by BLAST from the S09 database available during this final phase and then plotting the alignments for the *supercontigs* results in Figure 13. This view demonstrates the lack of detected alignments to the N-terminus. Also Figure 13 illustrates not just the maximal length but does highlight in yellow the substantial gaps within the stretch. Not only are there gaps within the *supercontig* coverage of the *pgp* but even with this latest database there are only two *pgp* fully covered by a single *supercontig* each.

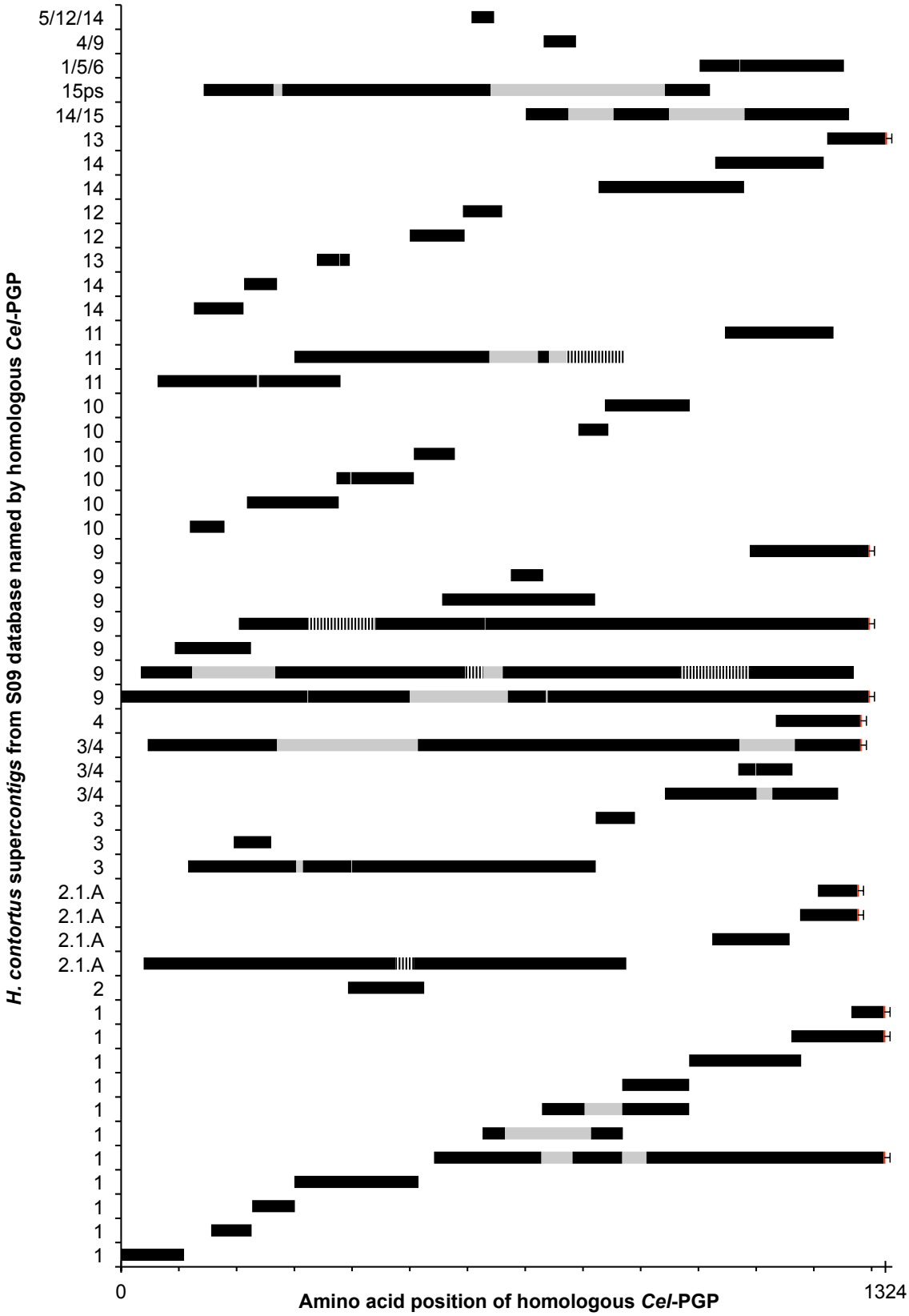


Figure 13: Alignments of *H. contortus supercontigs* from S09 Database with Homologous *Cel*-PGP.

Scale on *x* axis in amino acids of the homologous *Cel*-PGP; hence the STOP codon at different locations. *2.1.A* is *Hco*-PGP-2.1.A. Solid black bars: alignment. Vertical bars: alternative splices. Solid grey bars: unknown sequence between *contigs* within *supercontigs*. Red bars with cap: stop codons. Order is first by homologous single *Cel*-PGP or subgroup and second by 5' position of alignment. Each row is a single *supercontig*; *supercontig* number is in Table 8 below.

The detail of the alignments in Figure 13 are summarised in Table 8 with just the earliest start and latest end of the multiple and alternative alignments to the amino acid positions in the homologous *Cel*-PGP. The rows are in order and so indicate the number of the *H. contortus supercontig* for the applicable bar in Figure 13.

Table 8: Summarised Alignments of *H. contortus supercontigs* from S09 Database with Homologous *Cel*-PGP.

Homologous *Cel*-PGP	*H. contortus supercontig*	Earliest Amino Acid Aligned	Last Amino Acid Aligned
1	0050550	1	109
1	0045305	157	226
1	0034154	228	301
1	0061621	301	515
1	0006074	543	1321
1	0032856	627	665
1	0011236	730	984
1	0044638	869	984
1	0046186	985	1178
1	0047549	1162	1321
1	0024971	1266	1321
2	0049649	394	525
2.1.A	0007774	40	875
2.1.A	0038725	1025	1158
2.1.A	0046285	1177	1275
2.1.A	0005974	1208	1275
3	0001706	117	822
3	0059287	196	260
3	0023983	823	890
3/4	0000863	943	1242
3/4	0001880	47	1280
3/4	0001880	1070	1099
4	0057179	1135	1280
9	0006925	205	1294
9	0006925	1	1294
9	0006925	557	821
9	0038098	35	1269
9	0046372	94	225
9	0046372	676	731
9	0016882	1090	1294
10	0035472	120	179
10	0025718	219	377
10	0064200	374	507
10	0038139	508	578
10	0037674	793	844
10	0037517	839	985
11	0024002	64	380
11	0024351	301	871
11	0035404	1047	1234
14	0069412	127	212
14	0004549	214	270
13	0059902	340	396
12	0021658	501	595
12	0050285	593	660
14	0015622	828	1079
14	0021166	1030	1217
13	0010288	1224	1324
14/15ps	0024244	702	1261
15ps	0005706	144	1020
1/5/6	0013014	1003	1252
4/9	0055664	733	788
5/12/14	0055790	608	646

Eleven stop codons were detected within sequences that showed homology to *Cel*-PGP and are indicated in Figure 13 as final capped red bars. However, several of these sequences were relatively short and multiple *supercontigs* were homologous to the same *Cel*-PGP (three with *Cel*-PGP-1 and three with *Cel*-PGP-9) without corresponding numbers of TMD and ABC. In multiple regions of multiple *Cel*-PGP there was more than one *H. contortus supercontig* aligned. The author concluded the evidence in Figure 13 supported a total of **11** *H. contortus pgp* genes as summarised in Table 9.

Table 9: Number of *pgp* Genes in *C. elegans* and *H. contortus.*

***Cel*-PGP or Subgroup**	**Number of Genes in *C. elegans***	**Number of Homologues in *H. contortus***	**Difference *H. contortus* vs. *C. elegans***
1	1	1	0
2	1	1	0
3/4	2	2	0
5/6/7/8	4	0	**-4**
9	1	3	**+2**
10	1	1	0
11	1	1	0
12/13/14/15ps	3 & *ps*	2	**-1** / -2
TOTAL	**14 &** *ps*	**11**	**-3 / -4**

The *pgp* gene families are compared between *C. elegans* and *H. contortus* in the table above highlighting the lack of any evidence of homologues to the *Cel-pgp-5*, *6*, *7*, *8*; the two additional *pgp-9* genes and the fewer genes homologous to the *Cel-pgp-12*, *13*, *14*, *15ps* subgroup with the bioinformatic evidence from this study.

4 DISCUSSION

This thesis is concerned with revealing the *pgp* genes in *H. contortus*. The aim of this research was to discover the incomplete and dispersed P-gp genetic sequences in the sequence databases and construct the putative genes. These could then be investigated in the future for their rôle in IVM resistance in *H. contortus* and potentially utilised in molecular tests for IVM resistance.

In the introduction *H. contortus* was described as a parasitic nematode of sheep and goats that feeds on blood in the stomach, which can lead to symptoms of anaemia, lethargy, weight loss and even death in acute infections. As it is one of the most prevalent and economically important parasites internationally,[80] its treatment and control is paramount. The importance of this species was the reason for this study on it.

Treatment and control of helminths in general and *H. contortus* in particular is dependent on anthelmintics as there are no practical alternatives,[5; 6] to the extent that anthelmintics constitute 40% of all livestock health expenditure.[7] Commercial introduction of IVM was in 1981[9] and has grown to become the most commonly used drug in livestock agriculture.[7; 9] The importance of this drug was the reason for this study being directed at its use.

The high dependence on anthelmintics [5; 6] in general and IVM[7; 9] in particular means resistance has serious outcomes. There is already widespread resistance to the older anthelmintics including narrow-spectrum and all three classic broad-spectrum classes of drug, which has increased the reliance on IVM.[10; 11; 12] The emergence of IVM resistance in *H. contortus* is therefore a serious threat to the control of the pathogen[11; 13; 14] to the extent of threatening the entire sheep industry in Australia.[13] The significance of IVM resistance in *H. contortus* was the motivation behind this study into this particular combination.

Presently IVM resistance can be identified and measured on the basis of changes in phenotype. However, these phenotypic tests are not sensitive or reliable enough to detect early selection of resistance but are restricted to being indicators of clinical resistance.[16; 17; 18] The genetic basis of IVM resistance in *H. contortus*

is not currently known,[6; 71] which precludes any monitoring of resistant genotypes prior to phenotypic changes. The desire for genotypic tests for the benefit of academic research of resistance development and agricultural application of resistance control is the background for this thesis.

As the *H. contortus* genome has not yet been completed, and therefore not annotated, genome-wide approaches are currently not available for determining genes or loci under IVM selection that might serve as signals for genotype changes and tests. Therefore the approach taken to date is the analysis of individual genes or gene families implicated in IVM resistance. One candidate for the genetic route of IVM resistance is the P-gp family. There has been overwhelming evidence of their rôle in anthelmintic activity[60; 61] and P-gps have been associated with ML resistance from population genetic studies;[31; 62] although this was not consistent.[63] In another parasitic nematode, selection on an ABC transporter (the superfamily that includes P-gps) was verified,[64; 65; 66; 67] and in both *H. contortus* and the other nematode IVM treatment led to decreased polymorphism of P-gps, a classic indication of selection.[31; 68; 69] This background suggested the P-gp gene family should be investigated in this current study.

The P-gp family is well studied in many pathogens due to their involvement in multiple drug resistance to the extent that some of these proteins are named Multiple Drug Resistant Protein. However, in *H. contortus* the phylogenetics and ontology of the family is not yet established and even the definitive number of genes is unknown. As already stated, the *H. contortus* genome has yet to be completed and there is therefore no annotation of genes available. Only one P-gp mRNA transcript has been fully sequenced[70] and the other members have only partial sequences submitted, which are a fraction of their full length.[63; 68] The gene sequences remain fragmented and dispersed in the sequence databases. The P-gps are very long, contain multiple repeats and the family consists of multiple members that are similar; hence they are not easily amenable to most bioinformatic approaches.

The *pgp* genes have a repeated pair of domains with a 5' TMD and ABC and a 3' TMD and ABD. These domains are therefore physically close on the chromosomes. However, duplicated genes are also physically close in *C. elegans*, for example: the two members in the sub-group of *Cel-pgp-3* and *4*;

the four members of the sub-group of *Cel-pgp-5*, *6*, *7* and *8* or the four members of the sub-group of *Cel-pgp-12*, *13*, *14* and the pseudogene *15ps*. Therefore a pair of domains on a single *H. contortus* genomic fragment could be the 5' and 3' domains of a single gene but may well be the 3' domain of one gene and the 5' domain of a second separate gene. As the region before the 5' domain as well as between the 5' and 3' domains is hypervariable, discerning whether a sequence is 5' or 3' is limited. Because the initial *Met* and exon can be a long intron distant from the subsequent exons, the signal from the start of a 5' domain is not necessarily strong. The stop signal is an absolute signal for a domain being 3' unless a potential splice site is present and the stop signal is intronic. A number of sequences did not continue to the 3' end of the domain or unknown nucleotides between *contigs* in the *supercontig* blocked the signal. In the case of sequences only containing one domain with no 3' coverage of the potential stop signal the domain could be 5' or 3' in a gene. The homologous *Cel*-PGP for sequences with bi-domain coverage and a clear 5' start or 3' stop signal was determined by full-length comparison to both domains of all members of the *Cel*-PGP. In this case the only question was which individual or sub-group was the *H. contortus* sequence homologous to. Those sequences with bi-domain coverage but without a clear 5' start or 3' stop signal were analysed by comparison of each domain separately to *Cel*-PGP. In those cases the domains were independently homologous to 5' one and 3' for the other. In no case was a 3' domain contiguous with a 5' domain downstream in the sequence. In conclusion, no bi-domain sequences showed evidence of covering adjacent genes but did indicate that a single gene was present.

In the scenario of sequences only containing one domain with no 3' coverage of the potential stop signal, the domain could be 5' or 3' in a gene and it could be one of several gene homologues. The 5' and 3' domains of *Cel*-PGP were provided separately for homology in order to differentiate between both 5' and 3' as well as gene homology.

The above points do not take into account confounding signals in the sequence. The level of sequence difference between 5' and 3' domains in individual genes is similar to that between 5' domains in different genes in the same sub-group and between 3' domains in different genes in the same sub-group in *C. elegans* and is considered to be the same situation for *H. contortus*. In this case only weak

homology to one of the domains in one of the genes can be expected and was indeed found.

In addition to clear sequence providing weakened evidence, polymorphism in the genome of nematodes, including exons, is high and in *H. contortus* is particularly high. The outcome is that a sequence with no clear signal to a particular homologue can be swayed in one direction or another by polymorphisms. Therefore the stated homologous gene and the domain being identified as 5' or 3' is more tentative due to polymorphism.

Despite these difficulties it is considered worthwhile researching the genetic sequences of the *pgp* family so that the sequences and expression can be compared between sensitive and resistant populations and thus to assess *pgp* involvement in IVM resistance. The purpose of the bioinformatics in this study is to discover novel genetic sequences in the incomplete *H. contortus* genome that map both from and to known P-gp sequences and can be constructed into putative *H. contortus pgp* genes. The methodology has involved a combination of brute force sequence alignment followed by advanced bioinformatic algorithms.

The first stage of the bioinformatic approach taken in this research is of querying the incomplete *H. contortus* genome, via the multiple databases becoming available over the course of the research, with known *Cel*-PGP. Examining both the distribution of the HSPS and the histogram labelled coverage in Figure 1 clearly shows the cluster around amino acid position 350-600 and from 1000 to the C-terminus, these are the two ABC and are highly conserved enabling highly scored alignments. This increases the sensitivity of this method to detect homologous sequences in these regions. In the same way that the level of nucleotide variability between genes and individuals confounds determination of genomic sequence it also restricts the usefulness of local alignments as in BLAST.

Despite the presence of polymorphisms the P-gps, in common with all proteins, contained functional residues that were highly conserved. In addition, whole regions have particular functions and the range of variation at each residue is limited if the function of that domain is to be retained and for there not to be a fitness and survival penalty. This was exploited in the alignment of genomic sequences to known domains. This approach was the most apt for the P-gp

family as synonymous SNPs were disregarded and non-synonymous SNPs were assessed according to the matrix of scores depending on the specific position of the variant amino acid. This analysis was considered to come closer to a test of function and so more reliable as an indicator of gene identity than the direct study of sequence. However, full resolution of the TMD was restricted due to the lower conservation in comparison to the ABC.

The sections of unknown sequence in the *supercontigs* are a handicap to determining the full-length nucleotide and amino acid sequence. Not only are there gaps within the *supercontig* coverage of the *pgp* family but even with the latest database there are only two *pgp* fully covered by a single *supercontig* each. Several of the aligned sequences were relatively short thus reducing the reliability of these alignments. The summary of the *pgp* gene family coverage in Figure 13 demonstrates the lack of detected alignments to the N-terminus. Multiple *supercontigs* were homologous to the same *Cel*-PGP (three with *Cel*-PGP-1 and three with *Cel*-PGP-9) without corresponding numbers of TMD and ABC, in particular towards the N-terminus. In multiple regions of multiple *Cel*-PGP there was more than one *H. contortus supercontig* aligned. Whilst alternative splicing has been evident within *supercontigs*, determining whether overlapping *supercontigs* are alternative splices has not been conclusive.

There are eight clades in *C. elegans* (*Cel-pgp-1*; *2*; *3*/*4*; *5*/*6*/*7*/*8*; *9*; *10*; *11*; *12*/*13*/*14*/*15*) of which seven have representatives in *H. contortus*; there are no homologues to the entire clade of four genes *Cel-pgp-5* to *8* inclusive. All the duplications in *C. elegans* are on the X chromosome but none of these groups has more than one homologue in *H. contortus*. Conversely autosomal *Cel-pgp-2* and *Cel-pgp-9* with evidence of multiple homologues in *H. contortus* are not duplicated in *C. elegans*.

Despite *H. contortus* diverging from *C. elegans* approximately 400 million years ago[81] the striking aspect of the phylogenetics of the *pgp* members is the consistency with which *H. contortus pgp* cluster alongside *C. elegans* rather than the two species providing two distinct branches from the common root. One example of a neighbour-joining tree is shown in Figure 7 and other phylogenetic approaches demonstrated the same pattern. This is characterised as a deeply-rooted tree and means all of the intervening evolution has maintained the

homology within and comparative branching between each sub-group. Genes from the same sub-groups but different species have more similarity than genes from different sub-groups in the same species. Analysing just one *H. contortus* sequence removed any confounding potential of another *H. contortus* sequence changing the branching statistics. As the purpose of this comparison was to determine the identity of the *H. contortus* sequence in terms of which member or sub-group of the P-gp family and whether 5' or 3' domain, the tree included both species of *H. contortus* and *C. elegans* but was not rooted as the evolutionary distance was not required. Furthermore the sub-groups of the P-gp family are very deep rooted so that sub-group conservation is stronger than species differentiation. The inclusion of the entire bi-domain P-gp gene family from an outgroup species was considered to be an unnecessary complication.

A more positive aspect to this pattern of evolution is that it appears the *pgp* do not mutate quickly or substantially. With the assumption that the selective pressure on farms does not significantly alter this past record a simple diagnostic test for a resistant genotype could be designed that is not rapidly made void by mutations and indels. A quantitative measure of IVM resistance due to P-gp mechanism could thus be calculated. A different approach is the analysis of this family for *pgp*-specific targets as the deeply-rooted evolutionary trees suggest that these would remain effective over time.

The early evidence points to ten *pgp* genes but later evidence in

Figure 8 is for nine *pgp* in *H. contortus* and the results generated from the GeneWise algorithm indicates the number of stop codons is eleven as shown in Figure 13. The *pgp* family has a relatively large membership of 11 (*H. contortus*) or 15 (*C. elegans*) that presumably reflects the common requirement for those xenobiotic functions provided. Conversely the *Cel-pgp-5~8* clade would be conjectured to provide functions not required by the parasite or non-homologous proteins cover those ontologies. The substantial number of P-gps in the parasitic nematode provokes concern regarding the prevention or control of IVM resistance via P-gp mechanisms as multiple proteins may be capable of contributing to resistance under selective pressure. This scenario is particularly likely when *H. contortus* is exposed to multiple drugs and classes of anthelmintic.

The *in silico* approach taken in this thesis needs to be a rough sketch that an *in vitro* approach can take and provide the full picture. The sequences provided in full in the Appendix could be used for the design of PCR primers. The successful amplifications would provide proof of the primers existing on contiguous genetic lengths and sequencing would confirm or expand upon the predicted sequences in terms of the primary question of whether the *contigs* and *supercontigs* in the genomic databases are indeed contiguous *in vitro*. In addition evidence for or against the exons and introns predicted by the GeneWise algorithm could be provided. When mRNA would be used the exonic and by comparison the intronic fragments would be shown. If using gDNA then longer amplicons including the introns would confirm simply that the sequences constructed *in silico* at Sanger were correct. It may be that the portion of sequence the primers are designed upon are present, however, the intervening sequence is partially or wholly different *in vitro*. Whilst the gDNA cannot show exon-intron boundaries, splicing or splicing alternatives the amplicons based on mRNA templates would indeed provide evidence of these essential and important processes.

It is particularly important to become aware of alternative splices early in genotype studies as quantitative studies such as quantitative real-time PCR (qPCR) may either be based on only one of the alternative mature mRNA products and be misrepresenting (probably under-representing) the quantity and rate of transcription of the gene at rest or in response to stimuli. The other aspect is that a difference in quantity and rate of production of the alternative splices would not be detected if there are not separately measured amplicons for each alternative. This would be the case in both scenarios of an amplicon bridging over alternative splice sites and therefore detecting all alternatives or a primer site being located in only one of the alternative splices and thus the other splices, if and when produced, would not be amplified or detected. There could be up-regulation of the gene, only one of the splices or a differential up-regulation with one of the splices having a greater change than the other. A more dramatic and immediately interesting aspect of alternative splice regulation is if up-regulation of one splice and down-regulation of the other splice was seen. Where an alternative splice is possible the detection of each and every splice should be measured simultaneously, within each experiment for that gene.

The alternative splice potential is important in sequence differences between sensitive and resistant populations to ensure no mutations or indels are missed by not sequencing all alternative splices. On detection of a difference of regulation of at least one splice of one gene, the region upstream of the gene should be sequenced for mutations or indels between the sensitive and resistant populations that may be correlated with the differential regulation. The sequence difference could, in addition to being correlated, be the cause via interaction with regulatory and transcription factors.

An example of expression causing resistance has been the insertion of a transposable element into the regulatory region of the cytochrome P450 gene *Cyp6g1* causing up-regulation of the transcript in *D. melanogaster* and resistance to multiple insecticides that has spread globally.[57] IVM selected strains were associated with P-gp overexpression in *H. contortus*.[70] The linker domain in a P-gp between the ABC domains interacts with *α* and *β*-tubulin. These tubulin molecules are possibly arranged in the microtubules during the interaction. A suggested flow of information occurs with the P-gp & tubulin interaction dependent and differing according to microtubule integrity. As such the organelle status is signalled to the P-gp molecules. A further suggestion is that this signal up-regulates P-gp production.[82] The explanation for such a cascade being selected is that lowered cell defence from microtubule compromise is counteracted by the defence mechanism of permeable glycoprotein pumps expelling toxins across the cell membrane.

Sex-dependent selection of *β*-tubulin genotypes by IVM treatment was discovered with females being selected but not males over 3 years in the human parasitic nematode *O. volvulus*.[20] However, selection on male *β*-tubulin genotypes did occur over 6 years. [69; 83] The explanation may be selection is initially of the females only but then selection occurs in their progeny and is detected over longer time scales with the year-long life cycle of *O. volvulus*. Sex-dependent mortality was detected also in *O. volvulus* with an estimated decrease in the female population of 27.5%, double that of an estimated 13.5% in male nematodes.[84] The effect of IVM was therefore greater against female worms and could exert greater selection pressure on the females. Sex-dependent expression was estimated to be 24.3 fold higher in females of a gene putatively identified as a GluCl in the human parasitic nematode *Brugia malayi*.[85] As GluCl are thought to

be the main target of IVM[86] in certain configurations this could be significant for the IVM mode of action and efficacy between the different sexes in *B. malayi*. The expression was not measured during IVM treatment. The action of IVM is speculated to be more efficacious in the presence of more molecular targets and IVM to have a more pronounced outcome in female *B. malayi*.

If sex-dependent expression, selection and mortality were consistently greater in females across the species; a scenario in *H. contortus* could be envisaged of higher innate expression of IVM molecular targets in females, thus increasing selection for heritable resistance mechanisms in the face of greater mortality. Sequence and expression studies should therefore distinguish between male and female subjects and report results for both separately.

Whilst the *in vitro* work outlined above could discover a genotype of *H. contortus* associated with an observed phenotype of IVM resistance, it may not be the ultimate cause. Not withstanding the use to pure science of determining causal factors, a correlation has a potential rôle as a molecular test of IVM resistance in a *H. contortus* population. Even if later research shows the test to be one of correlation with a different causal factor it can be utilised in the meantime.

A clear use of a genetic test would be in addressing the acknowledged deficit in our knowledge and understanding of the factors necessary or advantageous for both the initial development and subsequent spread in usual agricultural practice[6]. Genotyping tests to a different drug were sensitive in *H. contortus* when treatment failure had not yet occurred,[87] were a similar standard of test developed for IVM resistant genotypes, the early stage of anthelmintic resistance could be detected and even subtle or slow effects could be followed. Cross-sectional studies of *H. contortus* may reveal variables correlated with IVM resistance and longitudinal research could follow the development of resistance in different flocks. The gold standard of blind randomised controlled trials could measure even slight changes to the *H. contortus* population genotype. This would contribute to a clear quantitative assessment of competing hypotheses for slowing, limiting or preventing resistance. These potentially useful chemotherapy and management regimes include treating according to clinical symptoms or treating less than 100% of the sheep with random allocation to treatment or non-treatment groups.

There is a very important advantage of examining genotypes, which is that a warning may be found that a phenotypically sensitive population has the genetic potential in at least some members to be IVM resistant. This pre-clinical resistance would be under selective pressure with traditional whole-flock prophylactic application of this broad-spectrum anthelmintic. Using a molecular test the use of anthelmintics could be adjusted to minimise or eliminate the selective pressure according to best practice at the time. A test that is informative before a problem exists is very helpful in the monitoring of *H. contortus* populations but it may be harder to market this approach. Particularly since the existing tests are not being fully utilised even when there is a need in flocks with suspicions of IVM resistance. Genetic tests will have to be understood to an even greater extent than the existing procedures.

It may be that such a genetic test would first only be able to be used at a specialised remote site with economies of scale. This organisation would be constrained by having to provide results in a time window that is still useful to the farm for making decisions in terms of chemotherapy and management. In terms of treating each animal or portion of a flock this may not be as practical and economically feasible as its use as a flock-wide surveillance method. The barriers to frequent use of a genetic test on many farms are predictable but worth stating. A practical test would have to be robust, sensitive and specific, simple, small, minimally invasive, use the minimum of equipment and not delicate expensive kit, use cheap consumables, be set up and run quickly and be able to be used by minimally trained and qualified personnel.

If such a molecular test could be made widely available then the surveillance and control of IVM resistance in *H. contortus* would be one step closer.

BIBLIOGRAPHY

1 Chappell, L. (1979). *Physiology of Parasites*. New York, U.S.A.: John Wiley.

2 Roberts, L., & Janovy, J. (2000). Foundations of Parasitology. New York, U.S.A.: McGraw Hill.

3 Blaxter, M. L. (2003). Nematoda: genes, genomes and the evolution of parasitism. *Adv. Parasitol.*, 54, pp. 101-195.

4 Hoberg, E. P., & Lichtenfels, J. R. (1994). Phylogenetic systematic analysis of the Trichostrongylidae (Nematoda), with an initial assessment of coevolution and biogeography. *J. Parasitol.*, 80(6) pp. 976-996.

5 Prichard, R.K., 2001. Genetic variability following selection of *Haemonchus contortus* with anthelmintics. *Trends Parasitol.*, 17 (9), pp. 445-53.

6 Wolstenholme, A.J., Fairweather, I., Prichard, R., von Samson-Himmelstjerna, G. and Sangster, N.C., 2004. Drug resistance in veterinary helminths. *Trends Parasitol.*, 20 (10), pp. 469-76.

7 Yates, D.M., Portillo, V. and Wolstenholme, A.J., 2003. The avermectin receptors of *Haemonchus contortus* and *Caenorhabditis elegans*. *Int. J. Parasitol.*, 33 (11), pp. 1183-93.

8 Chabala, J.C., Mrozik, H., Tolman, R.L., Eskola, P., Lusi, A., Peterson, L.H., Woods, M.F., Fisher, M.H., Campbell, W.C., Egerton, J.R. and Ostlind, D.A., 1980. Ivermectin, a new broad-spectrum antiparasitic agent. *J. Med. Chem.*, 23 (10), pp. 1134-6.

9 Campbell, W.C., 1989. *Ivermectin and abamectin*. New York: Springer-Verlag.

10 Bacon, J.A., Ulrich, R.G., Davis, J.P., Thomas, E.M., Johnson, S.S., Conder, G.A., Sangster, N.C., Rothwell, J.T., McCracken, R.O., Lee, B.H., Clothier, M.F., Geary, T.G. and Thompson, D.P., 1998. Comparative *in vitro* effects of closantel and selected β-ketoamide anthelmintics on a gastrointestinal nematode and vertebrate liver cells. *J. Vet. Pharmacol. Ther.*, 21 (3), pp. 190-8.

11 Jackson, F. and Coop, R.L., 2000. The development of anthelmintic resistance in sheep nematodes. *Parasitology*, 120 Suppl, pp. S95-107.

12 van Wyk, J.A. and Malan, F.S., 1988. Resistance of field strains of *Haemonchus contortus* to ivermectin, closantel, rafoxanide and the benzimidazoles in South Africa. *Vet. Rec.*, 123 (9), pp. 226-8.

13 Besier, R.B. and Love, S.C.J., 2003. Anthelmintic resistance in sheep nematodes in Australia: the need for new approaches. *Aust. J. Exp. Agr.*, 43 (12), pp. 1383-91.

14 Waller, P.J., 1997. Anthelmintic resistance. *Vet. Parasitol.*, 72 (3-4), pp. 391-412.

15 Coles, G.C., Bauer, C., Borgsteede, F.H., Geerts, S., Klei, T.R., Taylor, M.A. and Waller, P.J., 1992. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. *Vet. Parasitol.*, 44 (1-2), pp. 35-44.

16 Craven, J., Bjørn, H., Barnes, E.H., Henriksen, S.A. and Nansen, P., 1999. A comparison of *in vitro* tests and a faecal egg count reduction test in detecting anthelmintic resistance in horse strongyles. *Vet. Parasitol.*, 85 (1), pp. 49-59.

17 Humbert, J.F., Cabaret, J., Elard, L., Leignel, V. and Silvestre, A., 2001. Molecular approaches to studying benzimidazole resistance in trichostrongylid nematode parasites of small ruminants. *Vet. Parasitol.*, 101 (3-4), pp. 405-14.

18 Silvestre, A. and Cabaret, J., 2002. Mutation in position 167 of isotype 1 β-tubulin gene of Trichostrongylid nematodes: role in benzimidazole resistance? *Mol. Biochem. Parasitol.*, 120 (2), pp. 297-300.

19 de Lourdes Mottier, M., & Prichard, R. K. (2008). Genetic analysis of a relationship between macrocyclic lactone and benzimidazole anthelmintic selection on Haemonchus contortus. *Pharmacogenet. Genomics*, 18(2), pp. 129-140.

20 Bourguinat, C., Pion, S. D., Kamgno, J., Gardon, J., Duke, B. O., Boussinesq, M., & Prichard, R. K. (2007). Genetic selection of low fertile Onchocerca volvulus by ivermectin treatment. *PLoS Negl. Trop. Dis.*, 1(1), e72.

21 Gilleard, J. S., & Beech, R. N. (2007). Population genetics of anthelmintic resistance in parasitic nematodes. *Parasitology*, 134(8), pp. 1133-1147.

22 Kwa, M. S., Veenstra, J. G., & Roos, M. H. (1993). Molecular characterisation of β-tubulin genes present in benzimidazole-resistant populations of Haemonchus contortus. *Mol. Biochem. Parasitol.*, 60(1), pp. 133-143.

23 Kwa, M. S., Veenstra, J. G., & Roos, M. H. (1994). Benzimidazole resistance in Haemonchus contortus is correlated with a conserved mutation at amino acid 200 in β-tubulin isotype 1. *Mol. Biochem. Parasitol.*, 63(2), pp. 299-303.

24 Bartley, D., Jackson, F., Coop, R. L., Jackson, E., Johnston, K., & Mitchell, G. B. (2001). Anthelmintic-resistant nematodes in sheep in Scotland. *Vet. Rec.*, 149(3), pp. 94-95.

25 Bartley, D. J., Donnan, A. A., Jackson, E., Sargison, N., Mitchell, G. B., & Jackson, F. (2006). A small scale survey of ivermectin resistance in sheep nematodes using the faecal egg count reduction test on samples collected from Scottish sheep. *Vet. Parasitol.*, 137(1), pp. 112-118.

26 Pomroy, W. E. (2006). Anthelmintic resistance in New Zealand: a perspective on recent findings and options for the future. *N. Z. Vet. J.*, 54(6), pp. 265-270.

27 Nadler, S. A. (1987). Genetic variability in endoparasitic helminths. *Parasitol. Today*, 3(5), pp. 154-155.

28 Blouin, M. S., Dame, J. B., Tarrant, C. A., & Courtney, C. H. (1992). Unusual population genetics of a parasitic nematode: mtDNA variation within and among populations. *Evolution*, 46, pp. 470-476.

29 Blouin, M. S., Yowell, C. A., Courtney, C. H., & Dame, J. B. (1995). Host movement and the genetic structure of populations of parasitic nematodes. *Genetics*, 141(3), pp. 1007-1014.

30 Anderson, T. J., Blouin, M. S., & Beech, R. N. (1998). Population biology of parasitic nematodes: applications of genetic markers. *Adv. Parasitol.*, 41, pp. 219-283.

31 Blackhall, W. J., Liu, H. Y., Xu, M., Prichard, R. K., & Beech, R. N. (1998). Selection at a P-glycoprotein gene in ivermectin- and moxidectin-selected strains of Haemonchus contortus. *Mol. Biochem. Parasitol.*, 95(2), pp. 193-201.

32 Blackhall, W. J., Pouliot, J. F., Prichard, R. K., & Beech, R. N. (1998). Haemonchus contortus: selection at a glutamate-gated chloride channel gene in ivermectin- and moxidectin-selected strains. *Exp. Parasitol.*, 90(1), pp. 42-48.

33 Blackhall, W. J., Prichard, R. K., & Beech, R. N. (2003). Selection at a γ-aminobutyric acid receptor gene in Haemonchus contortus resistant to avermectins/milbemycins. *Mol. Biochem. Parasitol.*, 131(2), pp. 137-145.

34 Braisher, T. L., Gemmell, N. J., Grenfell, B. T., & Amos, W. (2004). Host isolation and patterns of genetic variability in three populations of Teladorsagia from sheep. *Int. J. Parasitol.*, 34(10), pp. 1197-1204.

35 Grillo, V., Jackson, F., & Gilleard, J. S. (2006). Characterisation of Teladorsagia circumcincta microsatellites and their development as population genetic markers. *Mol. Biochem. Parasitol.*, 148(2), pp. 181-189.

36 Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*, 97(1), pp. 145-163.

37 Thomas, W. K., & Wilson, A. C. (1991). Mode and tempo of molecular evolution in the nematode Caenorhabditis: cytochrome oxidase II and calmodulin sequences. *Genetics*, 128(2), pp. 269-279.

38 Urquhart, G.M., Armour, J., Duncan, J.L., Dunn, A.M., & Jennings, F.W. (1996). *Veterinary Parasitology*, 2nd ed. Oxford, U.K.: Blackwell Science.

39 Blouin, M. S. (1998). Mitochondrial DNA diversity in nematodes. *J. Helminthol.*, 72(4), pp. 285-289.

40 Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics*, 159(4), pp. 1805-1817.

41 Maynard Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.*, 23(1), pp. 23-35.

42 Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4), pp. 2335-2352.

43 ffrench-Constant, R. H., Daborn, P. J., & Le Goff, G. (2004). The genetics and genomics of insecticide resistance. *Trends Genet.*, 20(3), pp. 163-170.

44 Nair, S., Williams, J. T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P. N., Guthmann, J. P., Smithuis, F. M., Hien, T. T., White, N. J., Nosten, F. & Anderson, T. J. (2003). A selective sweep driven by pyrimethamine treatment in Southeast Asian malaria parasites. *Mol. Biol. Evol.*, 20(9), pp. 1526-1536.

45 Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A. C., Krishna, S., Nosten, F. & Anderson, T. J. (2007). Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.*, 24(2), pp. 562-573.

46 Kohn, M. H., Pelz, H. J., & Wayne, R. K. (2000). Natural selection mapping of the warfarin-resistance gene. *Proc. Natl. Acad. Sci. U.S.A.*, 97(14), pp. 7911-7915.

47 Silvestre, A., & Humbert, J. F. (2002). Diversity of benzimidazole-resistance alleles in populations of small ruminant parasites. *Int. J. Parasitol.*, 32(7), pp. 921-928.

48 Pennings, P. S., & Hermisson, J. (2006). Soft sweeps II - molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.*, 23(5), pp. 1076-1084.

49 Pennings, P. S., & Hermisson, J. (2006). Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics*, 2(12), e186.

50 Blouin, M. S. (2000). Neutrality tests on mtDNA: unusual results from nematodes. *J. Hered.*, 91(2), pp. 156-158.

51 Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L., & Thomas, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of Caenorhabditis elegans. *Science*, 289(5488), pp. 2342-2344.

52 Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M., & Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. *Nat. Genet.*, 37(5), pp. 544-548.

53 Denver, D. R., Morris, K., Lynch, M., & Thomas, W. K. (2004). High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. *Nature*, 430(7000), pp. 679-682.

54 Coles, G. C., Rhodes, A. C., & Wolstenholme, A. J. (2005). Rapid selection for ivermectin resistance in Haemonchus contortus. *Vet. Parasitol.*, 129(3), pp. 345-347.

55 Cabaret, J., & Gasnier, N. (1994). Farm history and breeding management influences on the intensity and specific diversity of nematode infection of dairy goats. *Vet. Parasitol.*, 53(3), pp. 219-232.

56 Grillo, V., Jackson, F., Cabaret, J., & Gilleard, J. S. (2007). Population genetic analysis of the ovine parasitic nematode Teladorsagia circumcincta and evidence for a cryptic species. *Int. J. Parasitol.*, 37(3), pp. 435-447.

57 Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., Feyereisen, R., Wilson, T. G., & ffrench-Constant, R. H. (2002). A single P450 allele associated with insecticide resistance in Drosophila. *Science*, 297(5590), pp. 2253-2256.

58 Coles, G. C. (1999). Anthelmintic resistance and the control of worms. *J. Med. Microbiol.*, 48(4), pp. 323-325.

59 Kaplan, R. M. (2004). Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol.*, 20(10), pp. 477-481.

60 Didier, A. and Loor, F., 1996. The abamectin derivative ivermectin is a potent P-glycoprotein inhibitor. *Anticancer Drugs*, 7 (7), pp. 745-51.

61 Pouliot, J.F., L'Heureux, F., Liu, Z., Prichard, R.K. and Georges, E., 1997. Reversal of P-glycoprotein-associated multidrug resistance by ivermectin. *Biochem. Pharmacol.*, 53 (1), pp. 17-25.

62 Kerboeuf, D., Blackhall, W., Kaminsky, R. and von Samson-Himmelstjerna, G., 2003. P-glycoprotein in helminths: function and perspectives for anthelmintic treatment and reversal of resistance. *Int. J. Antimicrob. Agents*, 22 (3), pp. 332-46.

63 Le Jambre, L.F., Lenane, I.J. and Wardrop, A.J., 1999. A hybridisation technique to identify anthelmintic resistance genes in *Haemonchus*. *Int. J. Parasitol.*, 29 (12), pp. 1979-85.

64 Keddie, E.M., Higazi, T. and Unnasch, T.R., 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Mol. Biochem. Parasitol.*, 95 (1), pp. 111-27.

65 Ardelli, B.F. and Prichard, R.K., 2004. Identification of variant ABC-transporter genes among Onchocerca volvulus collected from ivermectin-treated and untreated patients in Ghana, West Africa. *Ann Trop Med Parasitol*, 98 (4), pp. 371-84.

66 Ardelli, B.F., Guerriero, S.B. and Prichard, R.K., 2006. Ivermectin imposes selection pressure on P-glycoprotein from *Onchocerca volvulus*: linkage disequilibrium and genotype diversity. *Parasitology*, 132 (Pt 3), pp. 375-86.

67 Ardelli, B.F., Guerriero, S.B. and Prichard, R.K., 2006. Characterization of a half-size ATP-binding cassette transporter gene which may be a useful marker for ivermectin selection in *Onchocerca volvulus*. *Mol. Biochem. Parasitol.*, 145 (1), pp. 94-100.

68 Sangster, N.C., Bannan, S.C., Weiss, A.S., Nulf, S.C., Klein, R.D. and Geary, T.G., 1999. *Haemonchus contortus*: sequence heterogeneity of internucleotide binding domains from P-glycoproteins and an association with avermectin/milbemycin resistance. *Exp. Parasitol.*, 91 (3), pp. 250-7.

69 Eng, J.K. and Prichard, R.K., 2005. A comparison of genetic polymorphism in populations of *Onchocerca volvulus* from untreated and ivermectin-treated patients. *Mol. Biochem. Parasitol.*, 142 (2), pp. 193-202.

70 Xu, M., Molento, M., Blackhall, W., Ribeiro, P., Beech, R., & Prichard, R. (1998). Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog. *Mol. Biochem. Parasitol.*, 91(2), pp. 327-335.

71 Hejmadi, M.V., Jagannathan, S., Delany, N.S., Coles, G.C. and Wolstenholme, A.J., 2000. L-glutamate binding sites of parasitic nematodes: an association with ivermectin resistance? *Parasitology*, 120 (5), pp. 535-45.

72 Gottesman, M.M., Hrycyna, C.A., Schoenlein, P.V., Germann, U.A. and Pastan, I., 1995. Genetic analysis of the multidrug transporter. *Annu. Rev. Genet.*, 29, pp. 607-49.

73 Gottesman, M.M. and Pastan, I., 1993. Biochemistry of multidrug resistance mediated by the multidrug transporter. *Annu. Rev. Biochem.*, 62, pp. 385-427.

74 Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T. and Thomas, W.K., 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature*, 392 (6671), pp. 71-5.

75 Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.*, 89(22), pp. 10915-10919.

76 Altschul, S. F., & Gish, W. (1996). Local alignment statistics. *Meth. Enzymol.*, 266(2), pp. 460-480.

77 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), pp. 403-410.

78 Karlin, S., & Altschul, S. F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U.S.A.*, 87(6), pp. 2264-2268.

79 Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Res.*, 14(5), pp. 988-995.

80 Coop, R.L. and Jackson, F., 2000. Gastrointestinal helminthosis. *In:* Martin, W.B. and Aitken, I.D., eds. *Diseases of sheep*. 3rd ed. Oxford: Blackwell Science, pp. 159-67.

81 Vanfleteren, J.R., Van de Peer, Y., Blaxter, M.L., Tweedie, S.A., Trotman, C., Lu, L., Van Hauwaert, M.L. and Moens, L., 1994. Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences. *Mol Phylogenet Evol*, 3 (2), pp. 92-101.

82 Georges, E. (2007). The P-glycoprotein (ABCB1) linker domain encodes high-affinity binding sequences to α-and β-tubulins. *Biochemistry*, 46(25), pp. 7337-7342.

83 Eng, J. K. L., Blackhall, W. J., Osei-Atweneboana, M. Y., Bourguinat, C., Galazzo, D., Beech, R. N., Unnasch, T.R., Awadzi, K., Lubega, G.W., & Prichard, R. K. (2006). Ivermectin selection on β-tubulin: Evidence in Onchocerca volvulus and Haemonchus contortus. *Mol. Biochem. Parasitol.*, 150(2), pp. 229-235.

84 Gardon, J., Boussinesq, M., Kamgno, J., Gardon-Wendel, N., & Duke, B. O. (2002). Effects of standard and high doses of ivermectin on adult worms of Onchocerca volvulus: a randomised controlled trial. *Lancet*, 360(9328), pp. 203-210.

85 Li, B. W., Rush, A. C., Tan, J., & Weil, G. J. (2004). Quantitative analysis of gender-regulated transcripts in the filarial nematode Brugia malayi by real-time RT-PCR. *Mol. Biochem. Parasitol.*, 137(2), pp. 329-337.

86 Dent, J. A., Smith, M. M., Vassilatis, D. K., & Avery, L. (2000). The genetics of ivermectin resistance in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.*, 97(6), pp. 2674-2679.

87 von Samson-Himmelstjerna, G., Walsh, T.K., Donnan, A.A., Carrière, S., Jackson, F., Skuce, P.J., Rohn, K. and Wolstenholme, A.J., 2009. Molecular detection of benzimidazole resistance in *Haemonchus contortus* using real-time PCR and pyrosequencing. *Parasitology*, 136 (3), pp. 349-58.

APPENDIX
TABLE OF CONTENTS

Appendix Sequence Legend:

Green highlights both the exons and introns predicted by the GeneWise algorithm.

Yellow highlights the unknown nucleotides between the contigs within the supercontig and the unknown amino acids due to the identity of the nucleotides being unknown within the supercontig or the GeneWise algorithm being unable to extract coding sequence.

Blue and purple highlight alternative splicing.

A.1 *HAECO*_S09_SUPERCONTIG_0050550 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1_K08E7_9_1-1321AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0050550___HAECO_S09_Supercontig_0050550
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623

H_S09_0050550__-1472 [1472 :  711]


CAEEL-PGP-1____     1 MLRNGSLRQSLRTLDSFSLAPEDVLKTAIKTVEDYEGDNIDSNGEIKIT
                      M         ++ L S+       L  +  TV  Y    I S
                      MDFPFYKSNLIQVLISY-------LCESTSTV--YVHSKISSVLRKDEK
H_S09_0050550__ -710 agtcttatatacgcatt       ctgtatag  tgctaatagtaagga
                      tatctaacattatttca       tgacccct  atacatcgttgaaaa
                      gtttccaatacatatat       ttgataat  tttgatatgagacaa


CAEEL-PGP-1____    50 RDAKEEVVNKVSIPQL                       YRYTTTLEKL
                        KEE   KV + QL                       +RY T  E
                      AVVKEEPAKKVPLLQL                       WRYATWSELA
H_S09_0050550__ -590 gggaggcgaagcctct                       tctgatagtg
                      cttaaaccaatcttatGTGAGTA  Intron 1   CAGgaccggatc
                      accggggaagggaagg<0-----[542  :  291]-0>gttaagtaga


CAEEL-PGP-1____    76 LLFIGTLVAVITGAGLPLMSILQGKVSQAFINEQ              109
                      LLF+G  V+++TGAGLPLMSI+QG+V+QAF+ E+
                      LLFVGIFVSLVTGAGLPLMSIIQGQVTQAFVKEE
H_S09_0050550__ -260 tttggatgtcgagggtctataacgcgacgtgagga              157
                      ttttgtttcttcgcgtcttcttagatcacttaaat
                      ggttacttatatattagggccaaaactattcgga


H_S09_0050550__ -156  GTACAAG  Intron                               1
                      <2-----[156  :     ]-2>
//

Gene 1
EXONS 710 157
  Exon 710 543 phase 0
  Exon 290 157 phase 0
//

>H_S09_0050550___HAECO_S09_Supercontig_0050550
_710-157bp_AA
_CAEEL-PGP-1_1-109AA

MDFPFYKSNLIQVLISYLCESTSTVYVHSKISSVLRKDEKAVVKEEPAKKVPLLQLWRYA
TWSELALLFVGIFVSLVTGAGLPLMSIIQGQVTQAFVKEE
//

>H_S09_0050550___HAECO_S09_Supercontig_0050550

atggattttcctttctacaaatcaaatttaatccaagttctaatttcatatctttgtgag
tcaacttcaacagtttatgttcattcgaaaatttcaagtgtgttaaggaaagacgaaaaa
gcagtcgtcaaggaggagccggcaaaaaaggtgccgctattacagttgtggcgttatgca
acatggagtgaattggcattgttgtttgttggaatctttgtttcacttgtaactggagct
ggtttaccgttgatgtccatcatacaaggacaagtcactcaagctttgtcaaggaggaa
at
//
```

A.2 *HAECO*_S09_SUPERCONTIG_0045305 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:      CAEEL-PGP-1____CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence     H_S09_0045305___HAECO_S09_Supercontig_0045305

H_S09_0045305__     1          Intron CAG
                    <0-----[     :  170]-0>


CAEEL-PGP-1____   157 VTCYLYVAEQMNNRLRREFVKSILRQEISWFDTNHSGTLATKLF
                      VTC+L V EQM+NR+RR+FVK+IL Q+ISWFD N+SGTLATKLF
                      VTCFLIVCEQMSNRIRRKFVKAILHQDISWFDKNNSGTLATKLF
H_S09_0045305__   171 gatttagtgcaaacaccatgagatccgatttgaaaagacgaact
                      tcgttttgaatgagtggattacttaaatcgtaaaaggctccatt
                      ggctgttcgggctatagatgactatattagctatccgagctaac


CAEEL-PGP-1____   201                D          NLERVKEGTGDKIGMAFQYLSQF
                                     D          N+ERVKEGTGDKIG+ FQY SQF
                                     D          NIERVKEGTGDKIGLIFQYTSQF
H_S09_0045305__   303 g          [gac]          aagcgaggaggaagcatctaact
                      aGTAAGTA  Intron 1   CAG atagtaagcgaatgtttaacgat
                       <2-----[305  :  532]-2>cttaccgagtttacgcccgtttat


CAEEL-PGP-1____   225 IT                                              226
                      +T
                      VT
H_S09_0045305__   603 gag                                             609
                      tc
                      ta
```

```
H_S09_0045305__   610 GTGAATT  Intron                                1071
                        <1-----[610  :      ]-1>
//

Gene 1
EXONS 171 609
  Exon 171 304 phase 0
  Exon 533 609 phase 2
//

Making a D in phase 2 intron


>H_S09_0045305___HAECO_S09_Supercontig_0045305
_171-609bp_AA
_CAEEL-PGP-1_157-226AA

VTCFLIVCEQMSNRIRRKFVKAILHQDISWFDKNNSGTLATKLFDNIERVKEGTGDKI
GLIFQYTSQFVT
//

>H_S09_0045305___HAECO_S09_Supercontig_0045305

gtgacgtgctttttgattgtttgcgagcagatgagcaatcgaattcgacggaaa
tttgtgaaagccattttacatcaagatatttcatggttcgataaaaataacagcgggaca
ctggccactaaactattcgacaatattgaacgcgtcaaggaagggactggtgataaaatc
gggctcatcttccagtatactagtcaatttgttacag
//
```

A.3 *HAECO*_S09_SUPERCONTIG_0034154 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence      H_S09_0034154___HAECO_S09_Supercontig_0034154

H_S09_0034154__-1037           Intron     TAAG
                                <1-----[     :  629]--1>


CAEEL-PGP-1____  228  FIVAFTHSWQLTLVMLAVTPIQALCGFAIAK
                      F++AFTHSW+LTLVMLAVTP+QA CGFAI+K
                      FVIAFTHSWKLTLVMLAVTPLQAFCGFAISK
H_S09_0034154__ -628  tgagtacatatacgatggacccgttgtgata
                     gtttctcaggatcttttctcctactggtctca
                     ttccatatcgggaaaggtggttaactacattg


CAEEL-PGP-1____  259                          SMSTFAIRETLRYAKAGKVVEETISS
                                              SMSTF I E ++YAKAG++VE+T+SS
                                              SMSTFTIAEAIKYAKAGRMVEQTVSS
H_S09_0034154__ -533                          tatataagggaatgaggaaggcagtt
                     GTGAGGG  Intron 1   CAGctcctctcactaacacggttaactcc
                     <0-----[533  :  313]-0>aggcctatgtcattacaagcaatgga


CAEEL-PGP-1____  285 IRTVVSLNGLRYELER                              301
                     IRTV SLNGLRYE+ER
                     IRTVCSLNGLRYEIER
H_S09_0034154__ -234 acagttcagcctgaga                              188
                     tgctgctagtgaatag
```

```
                     tcggctctttccata


H_S09_0034154__ -187 GTATCTA  Intron                                 1
                     <2-----[187  :     ]-2>

//

Gene 1
EXONS 628 188
  Exon 628 534 phase 0
  Exon 312 188 phase 0
//

>H_S09_0034154___HAECO_S09_Supercontig_0034154
_628-188bp_AA
_CAEEL-PGP-1_228-301AA

FVIAFTHSWKLTLVMLAVTPLQAFCGFAISKSMSTFTIAEAIKYAKAGRMVEQTVSSIR
TVCSLNGLRYEIER
//

>H_S09_0034154___HAECO_S09_Supercontig_0034154

gttttgtcatcgcatttacacatagctggaagttgacactagtaatgttggctgtgacg
cctcttcaagcattctgtggattcgcaatttctaagtcaatgtcgaccttcactatagct
gaggctatcaaatatgctaaagccggaagaatggtcgaacaaactgtgtcgtcaattcgc
acggtgtgctctctcaatggtcttcgctacgaaattgaaag
//
```

A.4 *HAECO*_S09_SUPERCONTIG_0061621 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:      CAEEL-PGP-1___CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence     H_S09_0061621___HAECO_S09_Supercontig_0061621

H_S09_0061621__     1          Intron     CAG
                      <2-----[     :   47]-2>


CAEEL-PGP-1____   301  YSTAVEEAKKAGVLKGLFLGISFGAMQASNFISFALAFYIGVGW
                       Y  A+ EA++AG+LK LF+G+SF  M  +NF SFALAFYIG+ W
                       YKVALLEARRAGILKSLFVGLSFALMGLTNFSSFALAFYIGITW
H_S09_0061621__    48  taggccggccggacaattggcttgtagcaattttgtgttagaat
                       aatcttacggcgttagtttgtctcttgtcatcctctctatgtcg
                       gcgttacaatttctggtacttggtaggtgctccactgacctacag


CAEEL-PGP-1____   345 VHDGSLNFGDMLT                          TFSSVMMGSMALG
                        DG L + D++T                          TF SVMMGS+ALG
                      TVDGQLELKDLMT                          TFFSVMMGSLALG
H_S09_0061621__   181 agggctgtagcaa                          atttgaagttgcg
                      ctagatataattcGTGAGAG  Intron 1   CAGcttctttgctctg
                      attaagggatgga<0-----[220  :  312]-0>acttaggcggcct


CAEEL-PGP-1____   371 LAGPQLAVLGTAQGAASGIYEVLDR                       K
                       AGPQ AVLG AQGAA+ IYEVLDR                       +
                      QAGPQFAVLGAAQGAAASIYEVLDR                       E
H_S09_0061621__   352 cggcctggcgggcgggaatggcga                        g
                      acgcatcttgccagcccgtaattagGTCAGCG  Intron 2   CAGa
                      ataagtcttatcaattctctacacg<0-----[427  :  490]-0>a


CAEEL-PGP-1____   397 PVIDSSSKAGRKDMKIKGDITVENVHFTYPSRPDVP
                      P IDS+S+ GR+D+ IKG+I V+NV F YPSRPDVP
                      PEIDSTSQEGRRDVDIKGNIVVKNVVFNYPSRPDVP
H_S09_0061621__   494 cgagtaacggccgggaagaaggaaggtatctacggc
                      cataccgaagggatatagatttaatttaaccgcatc
                      tactttcaataacttcagttgggttccttccactaa


CAEEL-PGP-1____   433                        ILRGMNLRVNAGQTVALVGSSGCGKS
                                             +L+ ++L VNAG+TVALVGSSGCGKS
                                             VLKNLSLTVNAGETVALVGSSGCGKS
H_S09_0061621__   602                        gcaactcagagggaggcggtagtgaa
                      GTAAGCC  Intron 3   CAGttaatctctacgactcttgcggggag
                      <0-----[602  :  695]-0>tcgctaccgcctaggtgaaactttat


CAEEL-PGP-1____   459 TIISLLLRYYDVLKGK                          ITIDGVDVRD
                      TI+SLLLRYY+VLKG+                          I IDGV + D
                      TIVSLLLRYYNVLKGE                          IIIDGVPISD
H_S09_0061621__   774 aagacccctagcagg                           aaagggcatg


                      cttgtttgaaattagaGTTGGTT  Intron 4   AAGtttagtctca
                      cactaactcccataag<0-----[822  :  895]-0>cactctgtcc


CAEEL-PGP-1____   485 INLEFLRKNVAVVSQEPALFNCTIEENISLG                 515
                       IN+++LR  +AVVSQEP LFNCTIEENI +G
                       INIQYLRNRIAVVSQEPILFNCTIEENIRVG
H_S09_0061621__   926 aaacttcacagggtcgcactataaggaacggc               1019
                       tataatgagtcttcaactttagctaaatgtg
                       cctataatatttaaatcgcctccggctggt


H_S09_0061621__  1020  GTTAAAA                                       1069
                        <-----[1020 :     ]->
//

Gene 1
Gene 48 1019
  Exon 48 219 phase 0
  Exon 313 426 phase 0
  Exon 491 601 phase 0
  Exon 696 821 phase 0
  Exon 896 1019 phase 0
//

>H_S09_0061621___HAECO_S09_Supercontig_0061621
_48-1019bp_AA
_CAEEL-PGP-1_301-515AA

YKVALLEARRAGILKSLFVGLSFALMGLTNFSSFALAFYIGITWTVDGQLELKDL
MTTFFSVMMGSLALGQAGPQFAVLGAAQGAAASIYEVLDREPEIDSTSQEGRRDVDIKGN
IVVKNVVFNYPSRPDVPVLKNLSLTVNAGETVALVGSSGCGKSTIVSLLLRYYNVLKGEI
IIDGVPISDINIQYLRNRIAVVSQEPILFNCTIEENIRVG
//

>H S09 0061621___HAECO_S09_Supercontig_0061621
_48-1019bp

gtacaaggttgctctactcgaagcacgtcgtgctggcattctgaag
agtttattcgttggtctgtcgtttgcattgatgggtctgaccaatttctcctcattcgct
ttggcattctacattggaatcacatggacagttgatggacaattggagttgaaagatctg
atgacaacattcttttctgtaatgatgggctcgttggccctcggtcaagctggaccacag
tttgccgttcttggagctgcccaaggagctgctgccagtatctatgaagtcctagacagg
gaacctgaaatcgattctactagccaagaaggtcgacgagacgttgatatcaaagggaat
attgtggtgaagaatgttgtcttcaattatcccctccagacccgatgtaccagtttctcaag
aacctttcactcaccgtgaacgccggtgaaacggtggctctggtaggatcaagcggttgt
ggtaaaagtaccatagtcagtctactactccgttactacaacgtacttaaaggagagatc
ataatcgatggcgttccgatttccgacatcaacattcaatatttacgaaatcgaattgct
gttgtttcacaagaacctatcctgttcaactgtaccatcgaggagaacattcgggtgggt
c
//
```

A.5 *HAECO*_S09_SUPERCONTIG_0006074 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1_K08E7_9_1-1321AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0006074___HAECO_S09_Supercontig_0006074
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623
```

```
H_S09_0006074__    1          Intron      CAG
                   <0-----[      :  617]-0>


CAEEL-PGP-1____  543 GYNTLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATSALDAESEG
                     G  T VGDRGTQLSGGQKQRIAIARALVR+PKILLLDEATSALDAESE
                     GLQTNVGDRGTQLSGGQKQRIAIARALVRDPKILLLDEATSALDAESEA
H_S09_0006074__  618 gccaagggcgacctggcaccagagcgcgcgcaacccgggaagtgggagg
                     gtacatgaggcatcggaaagtctcgcttgacattttaaccgctacagac
                     tcgatgtttataaaaaggaaagtaagccttcgtcatcgaaccgtcacag


CAEEL-PGP-1____  592 IVQQALDK                          AAKGRTTIIIAHRLSTIR
                     +VQ+AL+K                          A+ GRTTIIIAHRLSTI+
                     VVQRALEK                          ASIGRTTIIIAHRLSTIK
H_S09_0006074__  765 ggccgcga                          gtagcaaaaagccctaaa
                     ttagctaaGTACAGT  Intron 1   CAGcctggcctttcagtccta
                     taagtgag<0-----[789  :  839]-0>agtttagtccccaggaca


CAEEL-PGP-1____  618 NADLIISCK                         NGQVVEVGDHRALMAQQ
                     NA  II+ K                          G+V EVG H  L+A +
                     NAHKIIAIK                         AGEVEEVGTHEELLANK
H_S09_0006074__  894 agcaaagaa                         gggggggacggccgaa
                     acaattctaGTTTGTT  Intron 2   TAGcgataatgcaaattcaa
                     ttcgctaca<0-----[921  : 1244]-0>ctgcgacttcgaagata


CAEEL-PGP-1____  644 GLYYDLVTAQTFTDAVDSAAE              G          KF
                     GLYYDLV AQTFTDAVD                  G           +
                     GLYYDLVHAQTFTDAVDDVNG              G          LL
H_S09_0006074__ 1296 gcttgcgcgcataggggggagg          [ggc]        cc
                     gtaaattacactcactaataggGTATTTC  Intron 3   CAG tt
                     tctctgacaagtcctcttcct <2-----[1361 : 2133]-2>cct


CAEEL-PGP-1____  668 SRENSVARQTSEHEG-----LSRQASEMDDIMNRVRSSTIGSITNGPVI
                     SRE SV      H         +  S    ++ RVR++TI ++    +
                     SRETSVQSANDLHVSWVSLPVEQCFSRTIRLQLRVRTATIMTMDEQQMD
H_S09_0006074__ 2141 tagatgctgagtcgatgtccggctttaaactccagcagaaaaaggccag
                     cgacctaccaatatggtctctaagtcgctgtatgtgcccttctaaaata
                     cagcatatcctgtgtgtatacaattaatttagcaaagactgcgtgaggt


CAEEL-PGP-1____  712 DEKEERIGKDALSRLKQ                          728
                     + KE+   KD ++RL++
                     ETKEK---KDDITRLRK
H_S09_0006074__ 2288 gaaga   aggaaccca                          2329
                     acaaa   aaatcgtga
                     gaaga   attttatta


CAEEL-PGP-1____  729              E          LEENNA             735
                                                +  + +
                                  A          ISTSMV
H_S09_0006074__2330 g          [gca]        atatag              2923
                     GTAAGCC  Intron 4   CAGctccctt
                     <1-----[2331 : 2904]-1>atcatg


CAEEL-PGP-1____  736    QKT                                     738
                        xxx
H_S09_0006074__2924     NNN                                     2933
                        NNN
                       NNNN
                       10 n [2924 : 2933]


CAEEL-PGP-1____  739 NLFEILY                                    745
                     N  EIL
                     N--EILL
H_S09_0006074__2934  a  qatt                                    2948
                     a  attt
                     t  aaaa


CAEEL-PGP-1____  746 H
                     !!!!
H_S09_0006074__ 2949 aatc                                       2952


CAEEL-PGP-1____  747 ARPHALS---LFIGMSTATIGGFIYPTYSVFFTSFMNVF    782
                     A P   S     I     IG     T    +F SF+N +
                     ASPSQKSSKTKGIRSHPVKIGLSYQITQITLFFSFQNCY
H_S09_0006074__ 2953 gtcacaataaagacaccgaagtatcaacaacttttcatt    3069
                     cccgaagcacagtggactatgtgaatcatctttctaaga
                     ttataatgacaattctctactgccgacgacgttatattt


                     !
H_S09_0006074__ 3070 a                                          3070


H_S09_0006074__ 3071 GTTTTTG  Intron 5   CAG
                     <0-----[3071 : 3306]-0>


CAEEL-PGP-1____  783 -AGNPADFLSQGHFWALMFLVLA
                        NP  +LS GHFWALMFLVLA
                     SFSNPDTLLSTGHFWALMFLVLA
H_S09_0006074__ 3307 tttacgatctagcttgcatcgcg
                     ctcacacttccgatgctttttc
                     gttcttagtagatcgtggctagc


CAEEL-PGP-1____  805 AAQGICSFLM                          TFFMGIASESLTRDLR
                       QG    F +                        TFFMG  +E+LT DLR
                     GIQGCTMPGQ                         TFFMGYGAENLTMDLR
H_S09_0006074__ 3376 gacgtaatgc                         attagtgggacaagtc
                     gtaggcttgaGTGAGAT  Intron 6   CAGctttgagcaatctatg
                     taaacagttg<0-----[3406 : 4034]-0>actgtctcactagtat


CAEEL-PGP-1____  831 NKLFRNVLSQHIGFFDSPQNASGKISTRLATDVPNLRT    868
                     +KLF N+LSQ +G+FDSP +ASGKI TRLATDVPNLR+
                     SKLFSNILSQDMGYFDSPLHASGKICTRLATDVPNLRS
H_S09_0006074__ 4083 tatttaactcgagttgtctcgagaatactgaggcatct    4196
                     cattcattcaatgatacctacggatgcgtccatcatgc
                     aagcgctcaacgtcttaaatctagataagtctattaaa


CAEEL-PGP-1____  869 AIDFRFSTVITT                             880
                        F    ++ T
```

```
                     VSIFNKDIMVLT
H_S09_0006074__ 4197 gaataagaagcat                            4233
                     tgttaaattttc
                     atatcgtcggtt


                      GTATCAT  Intron 7
                      <1-----[4234 :     ]-1>
                      304 n [5063 : 5366]


CAEEL-PGP-1____  881 LVSMVAGIGLAFFYGWQMALLIIAILPIVA            910
H_S09_0006074__      ------------------------------


H_S09_0006074__             Intron 8   CAG
                     <1-----[     : 5428]-1>


CAEEL-PGP-1____  911 FGQYLRGRRFTGK
                      GQ LR R + GK
                     IGQALRVRVMGGK
H_S09_0006074__ 5429 agcgtcgcgagga
                     gtgactgtgttgga
                     acaataatttgtcg


CAEEL-PGP-1____  924 NVKSASEFADSGK                            IAIEAIENVRTVQ
                     + K+A +F DSGK                          +A+EAIE+VRTVQ
                     HRKNAKDFEDSGK                          VAMEAIEHVRTVQ
H_S09_0006074__ 5470 ccaagagtggtga                          ggaggagcgcagc
                     agaacaataacgaGTACGAA  Intron 9   CAGtctactaatgcta
                     tagtaactattag<0-----[5509 : 5649]-0>atgaatgtttaca


CAEEL-PGP-1____  950 ALAREDTFYENFCEKLDIPHKEAIKEAFIQGLSY
                     AL +E+ F++ FC+ LD PH++A++E+FIQG++Y
                     ALTKEEAFHQKFCDYLDAPHRDALRESFIQGVAY
H_S09_0006074__ 5689 gcaagggtccattgtcggcccggccgttacgggt
                     ctcaaactaaatgaataccagactgacttagtca
                     tttggaactaatcttgtcatccaacaaccaaatt


CAEEL-PGP-1____  984                G          CASSVLYLLNTCAYRMGLALIIT
                                    G           A+S++Y+LN C+YR+GL LI++
                                    G          FATSIVYVLNCCSYRLGLYLIVS
H_S09_0006074__ 5791 g           [gga]        tgatagtgcattttctgctcaga
                     gGTTTGTT  Intron 10  TAG tcccttattaggcagtgtatttg
                      <2-----[5793 : 6063]-2>attatagtagcccatggtctgtcc


CAEEL-PGP-1____ 1008 DPPTMQPMRVL              R          VMYAITISTSTL
                        M P RVL               R          VMYAITIS+STL
                     S--IMMPTRVL              R          VMYAITISSSTL
H_S09_0006074__ 6134 a  aaacaagta           [aga]        gatgaaatttac
                     g  tttccgttgGTGAGTT  Intron 11  TAG ttactctcccct
                     t  tggagatg <2-----[6163 : 6220]-2>aagcccatgaatt


CAEEL-PGP-1____ 1032 GFATSYFPEYAKATFAGGIIFGMLRKISKIDSLSLAGEKK
                     GFA++YFPEY KA FAGGIIF ML++ S ID+L+  G+K+
                     GFASAYFPEYMKAAFAGGIIFNMLKQKSSIDNLTHDGKKE
H_S09_0006074__ 6258 gtgtgttcgtaaggtgggaataacacattagacacggaag
                     gtcccatcaatacctcggtttattaaacctaatcaagaaa
                     accgctctgtggatctctctccgcggacactccccccaag


CAEEL-PGP-1____ 1072                    KLYGKVIFKNVRFAYPERPEIEILK
                                        KL G + FKNV+F+YPERP+IE+LK
                                        KLSGAITFKNVKFSYPERPQIEVLK
H_S09_0006074__ 6378                    acaggaataagatatcgaccaggta
                     GTTTGAT  Intron 12  CAGatggctctaatatgacagcatatta
                     <0-----[6378 : 6447]-0>aacttctcatgacccgggtacaagg


CAEEL-PGP-1____ 1097                    GLSFSVEPGQTLALVGPSGCG
                                        GLSF+ +PG+TLALVGPSGCG
                                        GLSFTAKPGETLALVGPSGCG
H_S09_0006074__ 6523                    gtttagacggatgcggctgtg
                     GTATGTG  Intron 13  CAGgtctccacgactcttgccggg
                     <0-----[6523 : 6632]-0>cgactgatcacgactcatttt


CAEEL-PGP-1____ 1118 KSTVVALLERFYDTLGGEI                       FIDGSEI
                     KSTVV+L+ERFYD   G++                       ++D  ++
                     KSTVVSLIERFYDVKAGQV                       LLDSHDL
H_S09_0006074__ 6696 aaaggtcagcttggaggcg                       ccgtcgc
                     agcttcttagtaatacgatGTAGGCG  Intron 14  CAGttacaat
                     acatcggtgatctcgtaag<0-----[6753 : 6807]-0>gctacct


CAEEL-PGP-1____ 1144 KTLNPEHTRSQIAIVSQEPTLFDCSIAENIIYGLD
                     ++LNP HTRSQIAIVSQEP LFDCSIA+NI+YG++
                     RSLNPYHTRSQIAIVSQEPILFDCSIADNIVYGME
H_S09_0006074__ 6829 ctcactcactcagagtcgcactgttaggaagtgag
                     gctacaacgcatcttcaactttagctcaattagta
                     tagtatcaaaattccagaataacctttccttacaga


CAEEL-PGP-1____ 1179                -          -PSSVTMAQVEEAARLANIHNFI
                                                PS     ++E AAR ANIH FI
                                    V          RPS---QTEIENAARKANIHTFI
H_S09_0006074__ 6934 g           [gta]        cct   cagagaggcagaacata
                      GTGATTA  Intron 15  CAGtgcc   acataaccgacatactt
                      <1-----[6935 : 7442]-1>atta   acacacctcattccgtc


CAEEL-PGP-1____ 1201 AELPE                          GFETRVGDRGTQLSGGQKQRI
                       LP+                          G+ET VGD+GTQLSGGQKQRI
                     KGLPD                          GYETGVGDKGTQLSGGQKQRI
H_S09_0006074__ 7505 agccg                          gtgaggggagacctggcacca
                     agtcaGTAAGAA  Intron 16  TAGgaacgtgaagcatcggaaagt
                     gaatt<0-----[7520 : 7576]-0>tcaacatcaatggtttggatc


CAEEL-PGP-1____ 1227 AIARALVRNPKILLLDEATSALDTESEK
                     AIARAL+R+PKILLLDEATSALDTESEK
                     AIARALIRSPKILLLDEATSALDTESEK
H_S09_0006074__ 7640 gagcgcaaacaacccgggaagcgagaga
                     ctcgcttggcattttaaccgctacagaaGTAACGC  Intron 17
                     tctatttactgaggatgctctctgatag<0-----[7724 : 7815]


CAEEL-PGP-1____ 1255    VVQEALDRAREGRTCIVIAHRLNTVMNADCIAVVSNGTIIEK
                        VVQEALDRAREGRTCIVIAHRL+TV+NADCIAVV  G IIE+
                        VVQEALDRAREGRTCIVIAHRLSTVVNADCIAVVKGGVIIEQ
H_S09_0006074__ 7813    ggcggtgcgcggcatagagcccaaggaggtagggagggaagc
                     CAGttaactagcgaggcgtttcagtgcttacagtcttaggtttaa
                     -0>gaaaagtattattatattcttcttcctgtcttcaactgctaa


CAEEL-PGP-1____ 1297                G          THTQLMSEKGAYYKLTQKQ
                                    G          THT+LM+++G YY+LTQKQ
                                    G          THTELMAKRGFYYELTQKQ
H_S09_0006074__ 7942 g           [gga]        acagcagacgtttgcacac
                      GTTCGGT  Intron 18  CAGgcacattcaggtaaatcaaa
                      <1-----[7943 : 8640]-1>attgatgaaatcccaccaga


CAEEL-PGP-1____ 1317 MTEKK                                      1321

                     TVKSASE
H_S09_0006074__ 8700 agatgag                                    8720
                     ctaccga
```

```
                    ttgtcta

CAEEL-PGP-1____     *
                    *
                    *
H_S09_0006074__ 8721 t                                          8723
                    a
                    a


H_S09_0006074__ 8724        [8724 :20566]                      20566
                      554 n [12468:13021]
                       10 n [15042:15051]
                       10 n [17142:17151]
//

Gene 1
EXONS 618 8720
  Exon 618 788 phase 0
  Exon 840 920 phase 0
  Exon 1245 1360 phase 0
  Exon 2134 2329
       2330 2330 phase 2
  Exon 2905 3070 phase 1
  Exon 3307 3405 phase 0
  Exon 4035 4196
       4197 4233 phase 0

  ---

  Exon 5429 5508 phase 1
  Exon 5650 5792 phase 0
  Exon 6064 6162 phase 2
  Exon 6221 6377 phase 2
  Exon 6448 6522 phase 0
  Exon 6633 6752 phase 0
  Exon 6808 6934 phase 0
  Exon 7443 7519 phase 1
  Exon 7577 7723 phase 0
  Exon 7816 7942 phase 0
  Exon 8641 8720 phase 1
//

Making a G in phase 2 intron

Making a A in phase 1 intron

Making a G in phase 2 intron
Making a R in phase 2 intron
Making a V in phase 1 intron
Making a G in phase 1 intron


>H_S09_0006074___HAECO_S09_Supercontig_0006074
_618-2329bp_AA
_2330-2923bp_AA
---
_2934-2948bp_AA
_2949-2952bp_Frameshift
_2953-3069bp_AA
_3070bp_Frameshift
_3307-4196bp_AA
_4197-4233bp_AA
---
_5429-8720bp_AA
_CAEEL-PGP-1_543-728AA_729-735AA---739-745AA_746AA_Frameshift_747-
782AA_Frameshift_783-868AA_869-880AA---911-1321AA_End

GLQTNVGDRGTQLSGGQKQRIAIARALVRDPKILLLDEATSALDAESEAVVQRALEKASI
GRTTIIIAHRLSTIKNAHKIIAIKAGEVEEVGTHEELLANKGLYYDLVHAQTFTDAVDDV
NGGLLSRETSVQSANDLHVSWVSLPVEQCFSRTIRLQLRVRTATIMTMDEQQMD
ETKEKK
DDITRLRK

AISTSMV

NEILL

ASPSQKSSKTKGIRSHPVKIGLSYQITQITLFFSFQNCY


SFSN
PDTLLSTGHFWALMFLVLAGIQGCTMFGQTFFMGYGAENLTMDLRSKLFSNIL
SQDMGYFDSPLHASGKICTRLATDVPNLRS

VSIFNKDIMVLT
```

```
IGQALRVRVMGGKHRKNAKDFEDSGKVAMEAIEHVRTVQALTKEEAFHQKFCDYLD
APHRDALRESFIQGVAYGFATSIVYVLNCCSYRLGLYLIVSSIMMPTRVLRVMYAITISS
STLGFASAYFPEYMKAAFAGGIIFNMLKQKSSIDNLTHDGKKEKLSGAITFKNVKFSYPE
RPQIEVLK
GLSFTAKPGETLALVGPSGCGKSTVVSLIERFYDVKAGQVLLDSHDL
RSLNPYHTRSQIAIVSQEPILFDCSIADNIVYGMEVRPSQTEIENAARKANIHTFIKGLP
DGYETGVGDKGTQLSGGQKQRIAIARALIRSPKILLLDEATSALDTESEKVVQEALDRAR
EGRTCIVIAHRLSTVVNADCIAVVKGGVIIEQGTHTELMAKRGFYYELTQKQ
TVKSASE
//

>H_S09_0006074___HAECO_S09_Supercontig_0006074
_618-2329bp
_2330-2923bp
---
_2934-2948bp
_2949-2952bp_Frameshift
_2953-3069bp
_3070bp_Frameshift
_3307-4196bp
_4197-4233bp
---
_5429-8720bp

ggtctccagacaaatgtgggtgatcgtggaactcaactatcaggaggacagaagcaacga
atagcgattgcacgagcgctcgtccgtgatcccaagattctcctacttgacgaggcaaca
agcgccttggatgccgaaagcgaagcggttgtacaacgggctctggaaaaggcatcgatt
ggtcgtacaacgattatcatcgcccaccgactgtcgacaatcaaaaatgctcacaagatc
attgcaatcaaagccggtgaggtcgaggaagtcggtactcacgaggaactactggcaaat
aaaggtctctattacgatctggtacacgcacaaacgtttaccgacgctgtcgatgatgtc
aacggtggcctcctttccagagagacctcagttcaatctgccaacgatttgcatgtgagt
tgggtttcacttccagtcgaacaatgtttttcaagaactattcgtttacagctcagagta
cgaacggcaaccattatgaccatggatgagcaacagatggat
gagacaaaagagaaaaaa
gatgatattactcgacttcgtaaa

gcaatttccacatctatggt


aatgaaatattatta

aatc
gcttctccaagtcaaaaaagttcgaaaaccaaaggaattcgtagccatcccgttaaaatc
ggtttgagctaccagataacccagataaccctgttttttttcatttcaaaattgttat

a

tcgttttct
aaccctgatacattgctttcaacgggacatttctgggctctgatgttccttgtactggcc
ggtatacaaggatgcacaatgtttggtcagacattctttatgggttacggtgccgaaaac
cttacaatggatttacgttcaaaattgttctcgaacattctctcacaagacatgggttac
tttgattcaccattacatgccagtggaaagatatgtacacgattggctaccgatgtacct
aatttacgatca

gtaagtatatttaacaaggatatcatggtgcttactt


gaatcgga
caagctttacgagttcgtgttatgggtggcaagcatcgaaagaatgcaaaagactttgaa
gattctggaaaggtagctatggaagcaattgagcatgttcgtacagtccaagctcttact
aaggaggaagcattccatcaaaaattttgcgattatctggatgccccacatcgcgacgca
ctacgcgaatcattcatccaaggagtagcttatggatttgctacatctatagtgtatgta
ctgaactgctgctcatatcggttgggtctctatctgattgtcagcagtattatgatgcca
acgagagttttgagagtaatgtacgccatcacaatttcgtcatcaactcttggattcgcc
tcggcctatttccctgagtatatgaaggcagctttcgctggcggtatcattttcaacatg
ctcaagcagaaatcctcaatcgataacctcacccacgacggcaaaaaagagaaactaagc
ggtgctatcactttcaaaaatgtgaaattcagctacccggagaggcctcaaatcgaagta
ttgaag
ggcttgtcattcactgcgaaacctggcgaaaccttggca
ctcgttggcccatctggttgtggtaaaagcacagttgtctcgctgattgagcgattttac
gatgtcaaggctggacaagtgctgctcgattcacacgaccttcgttcactgaatccatat
cacacacgatcacaaattgctatcgtctcacaggaaccaattctattcgactgttctatt
gccgacaatattgtatacggaatggaagtacgtccttcacaaaccgaaatcgaaaacgcc
gctcgcaaagctaatatccacacgtttatcaagggactacctgatggttacgaaacaggc
gtaggtgacaaaggaactcagctgtctggtggtcagaagcaacgtatcgctatcgctcga
gctcttattagaagccctaagatactgctgctagatgaggccactagcgctctcgatacg
gaaagtgaaaaggtggtacaagaagcattggatcgagctcgtgaaggtcgtacatgtata
gttattgcccatcgtctcagtactgtcgtcaatgcggattgcattgctgtcgtaaaaggc
ggtgtgatcattgaacaaggaactcatacggaacttatggcaaaacgaggtttctactac
gaactcacccaaaagcaa
actgttaagtctgccagtgaa
//
```

A.6 *HAECO*_S09_SUPERCONTIG_0032856 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1 K08E7 9 1-814AA
Target Sequence      H_S09_0032856___HAECO_S09_Supercontig_0032856

H_S09_0032856__-1806          Intron     TAG
                     <0-----[     : 1327]-0>


CAEEL-PGP-1____   627 NGQVVEVGDHRALMAQQGLYYDLVTAQTFTDAVDSAAEG          665
                        G+V EVG H  L+A +GLYYDLV AQTFTDAVD    G
                      AGEVEEVGTHDELLANKGLYYDLVHAQTFTDAVDDVNGG
H_S09_0032856__-1328 ggggggggacggccgaagcttgcgcgcataggggggagg        1213
                      cgataatgcaaattcaagtaaattacactcactaatagg
                      ctgcgacttctaagatatctctgacaagttctcttcct

H_S09_0032856__-      GTATTTC
                      <2-----[1212 :     ]-2>
                        10 n [1062 : 1053]


CAEEL-PGP-1____   666 KFSRENSVARQTSEHEGLSRQASEMDDIMNRVRSSTIGSITNGPVIDEKE
H_S09_0032856__-      ----------------------------------------------------


CAEEL-PGP-1____   716 ERIGKDALSRLKQELEENNAQKTNLFEILYHARPHALSLFIGMSTATIGG
H_S09_0032856__-      ----------------------------------------------------


CAEEL-PGP-1____   766 FIYPTYSVFFTSFMNVFAGNPADFLSQGHFWALMFLVLAAAQGICS
H_S09_0032856__-      ---------------------------------------------


CAEEL-PGP-1____   812 FLM                                              814
H_S09_0032856__-      ---
//

Gene 1
EXONS 1328 1213
  Exon 1328 1213 phase 0
//

>H_S09_0032856___HAECO_S09_Supercontig_0032856
_1328-1213bp_AA
_CAEEL-PGP-1_627-665AA

A
GEVEEVGTHDELLANKGLYYDLVHAQTFTDAVDDVNGG
//

>H_S09_0032856___HAECO_S09_Supercontig_0032856
gcc
ggtgaggtcgaggaagtcggtactcacgatgaactactggcaaataaaggtctctattac
gatctggtacacgcacaaacgtttactgacgctgtcgatgatgtcaacggtgg
//
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1_K08E7_9_1-1321AA
Target Sequence      H_S09_0032856___HAECO_S09_Supercontig_0032856
H_S09_0032856__ -            Intron     CAG
                     <0-----[     :  856]-0>


CAEEL-PGP-1____   815 TFFMGIASESLTRDLRNKLFRNVLSQHIGFFDSPQNASGKISTRLATD
                        TFFMG  +E+LT DLR+KLF N+LSQ++G+FDSP +ASGKI TRLATD
                        TFFMGYGAENLTMDLRSKLFSNILSQNMGYFDSPLHASGKICTRLATD
H_S09_0032856__ -855 attagtgggacaagtctatttaactcaagttgtctcgagaatactgag
                        ctttgagcaatctatgcattcattcaatgatacctacggatgcgtcca
                        gccgtctcactagtatgagcgctcaacgtcttaaatctagctaagtct


CAEEL-PGP-1____   863 VPNLRTA                                          869
                        VPNLR+
                        VPNLRSV
H_S09_0032856__ -711 gcatctga                                         690
                        tcatgct
                        attaaga


H_S09_0032856__ -689  GTATATT  Intron                                   1
                      <1-----[689  :     ]
//

Gene 1
EXONS 855 690
  Exon 855 690 phase 0
//

>H_S09_0032856___HAECO_S09_Supercontig_0032856
_855-690bp_AA
_CAEEL-PGP-1_815-869AA

TFFMGYGAENLTMDLRSKLFSNILSQNMGYFDSPLHASGKICTRLATDVPNLRSV
//

>H_S09_0032856___HAECO_S09_Supercontig_0032856

acgttcttcatgggttacggtgccgaaaaccttacaatggatttacgttcgaaattg
ttctcgaacattctctcacaaaaacatgggttactttgattcaccattacatgccagtgga
aagatctgtacacgattggctaccgatgtacctaatttacgatcggtaa
//
```

A.7 *HAECO*_S09_SUPERCONTIG_0011236 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:      CAEEL-PGP-1____CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence     H_S09_0011236___HAECO_S09_Supercontig_0011236

H_S09_0011236__    1          Intron     CAG                          55
                   <2-----[     :   55]-2>


CAEEL-PGP-1____  730  LEENNAQKTNLFEILYHARPHALSLFIGMSTATIGGFIYPTYSVFFTS
                      L+    +KTNL+EIL +A PH   L +G++   IGG++YPTYSV F
                      LKAEGVKKTNLLEILAYASPHWKMLTVGLTACVIGGLVYPTYSVVFMQ
H_S09_0011236__   56  taggggaaaactgacgtgacctaataggcagtgaggcgtcattggtac
                   atacagtaacattattcacgcagattctgtccgttggttaccactttta
                   aggtaagggatgaatcttctctgagaggacaactcagtctaatcgtcga


CAEEL-PGP-1____  778 FMN                        VFAGNPADFLSQGHFWALM
                      +                         F+ NP  +LS GHFWALM
                     VIT                       SFS-NPDTLLSTGHFWALM
H_S09_0011236__  202 gaa                       ttt acgatctagcttgca
                     ttcGTAAGGA  Intron 1   CAGctc acacttccgatgctt
                     gtg<0-----[211  :  696]-0>att cttagtagatcgcgg


CAEEL-PGP-1____  800 FLVL                                          803
                     FLVL
                     FLVL
H_S09_0011236__  751 tcgc                                          762
                     tttt
                     ctag
                          13 n [1319 : 1331]


CAEEL-PGP-1____  804 AAAQGICSFLMTFFMGIASESLTRDLRNKLFRNVLSQHIGFFDSPQNASG
H_S09_0011236__      ------------------------------------------------


CAEEL-PGP-1____  854 KISTRLATDVPNLRT                              868
H_S09_0011236__      ---------------


H_S09_0011236__          Intron 3   TAG
                         <0-----[     : 2265]-0>


CAEEL-PGP-1____  869 AIDFRFSTVITTL
                     AIDFR+STVI TL
                     AIDFRLSTVIMTL
H_S09_0011236__ 2266 gagtctaagaaat
                     ctatgtgctttct
                     tctctgtgccgtg


CAEEL-PGP-1____  882 VSMVAGIGLAFFYGWQMALLIIAILPIV              A
                     +SM+AGI LAFFYGWQMA L++ ILP++
                     ISMLAGIVLAFFYGWQMAFLVVGILPLL             G
H_S09_0011236__ 2305 atatggagtgtttgtcagttgggacctcg           [gga]
                     tcttcgtttcttaggatcttttgttctt GTAAAAG  Intron 4
                     tcgatttagactttgagatgttatttat <1-----[2390 : 2469]


CAEEL-PGP-1____  910     FGQYLRGRRFTGKNVKSASEFADSGK
                         +GQ LR R + GK+ K+A +F DSGK
                         IGQALRVRVMGGKHRKNAKDFEDSGK
H_S09_0011236__ 2467     agcgtcgcgaggaccaagagtggtga
                     CAGgtgactgtgttggaagaacaataacga
                     -1>acaataatttgtcgtagtaactattag
```

```
CAEEL-PGP-1____  937                          IAIEAIENVRTVQALAREDTFYENFC
                                              +A+EAIE+VRTVQAL +E+ F++ FC
                                              VAMEAIEHVRTVQALTKEEAFHQKFC
H_S09_0011236__ 2550                          ggaggagcgcagcgcaagggtccatt
                     GTACGAA  Intron 5   CAGtctactaatgctactcaaactaaatg
                     <0-----[2550 : 2690]-0>atgaatgtttacactcggaactaacc


CAEEL-PGP-1____  963 EKLDIPHKEAIKEAFIQGLSYG                        984
                     + LD PH++A++E+FIQG++YG
                     DYLDAPHRDALRESFIQGVAYG
H_S09_0011236__ 2769 gtcggcccggccgttacgggtg                       2833
                     aataccagactgacttagtcag
                     ttactatccattaaccaaatt


H_S09_0011236__ 2834 GTTTGTT  Intron                              3031
                     <2-----[2834 :     ]-2>
//

Gene 1
EXONS 56 2833
  Exon 56 210 phase 0
  Exon 697 762 phase 0
  Exon 2266 2389 phase 0
  Exon 2470 2549 phase 1
  Exon 2691 2833 phase 0
//

Making a G in phase 1 intron


>H_S09_0011236___HAECO_S09_Supercontig_0011236
_56-762bp_AA---2266-2833bp_AA
_CAEEL-PGP-1_730-803AA---869-984AA

LKAEGVKKTNLLEILAYASPHWKMLTVGLTACVIGGLVYPTYSVVFMQVITSFSNPDTL
LSTGHFWALMFLVL


AIDFRLSTVIMTLISMLAGIVLAFFYGWQMAFLVVGILP
LLGIGQALRVRVMGGKHRKNAKDFEDSGKVAMEAIEHVRTVQALTKEEAFHQKFCDYLDA
PHRDALRESFIQGVAYG
//

>H_S09_0011236___HAECO_S09_Supercontig_0011236

aattgaaggctgaaggagtgaagaagacaaatctgttagaaattctcgcttatgccagt
ccccattggaaaatgttaacggtgggactcacagcatgcgttatcggagggcttgtctat
ccaacatattccgtggttttcatgcaagtgattacgtcattttctaaccctgatacattg
ctttcaacgggacatttctgggccctgatgttccttgtactg


gctatcgatttccgtttgagtacggtcatcatgactttgatttccatgttagctggtatt
gtattggcattcttttatggttggcaaatggcatttttggttgttggaattcttccttta
cttggaatcggacaagctttacgagttcgtgttatgggtggcaagcatcgaaagaatgca
aaagactttgaagattctggaaaggtagctatggaagcaattgagcatgttcgtacagtc
caagcccttaccaaggaggaagcattccatcaaaaattctgcgattatctagacgctcca
catcgcgacgcacttcgtgaatcattcatccaaggagtagcttatgg

//
```

A.8 *HAECO*_S09_SUPERCONTIG_0044638 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1___CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence      H_S09_0044638___HAECO_S09_Supercontig_0044638

H_S09_0044638__    1          Intron     TAG
                              <0-----[     :  376]-0>


CAEEL-PGP-1____  869 AIDFRFSTVITTLVSMVAGIGLAFFYGWQMALLIIAILPIV
                     AIDFR+STVI TL+SM+AGI LAFFYGWQMA L++ ILP++
                     AIDFRLSTVIMTLISMLAGIVLAFFYGWQMAFLVVGILPLL
H_S09_0044638__  377 gagtctaagaaatatatggagtgtttgtcagttgggacctc
                     ctatgtgctttcttcttcgtttcttaggatcttttgttctt
                     actttgtgccgagtagatttagaccttgagatgttatttat


CAEEL-PGP-1____  910                A          FGQYLRGRRFTGKNVKSASEFAD
                                               +GQ LR R + GK+ K+A +F D
                                    G          IGQALRVRVMGGKHRKNAKDFED
H_S09_0044638__  500 g          [gqa]          agcgtcgcgaggaccaagagtgg
                      GTAAAAG  Intron 1   CAGgtgactgtgttggaagaacaataa
                      <1-----[501  :  580]-1>acaataatttgtcgtagtaactat


CAEEL-PGP-1____  934 SGK                         IAIEAIENVRTVQALAREDTFYE
                     SGK                         +A+EAIE+VRTVQAL +E+ F++
                     SGK                         VAMEAIEHVRTVQALTKEEAFHQ
H_S09_0044638__  652 tga                         ggaggagcgcagcgcaagggtcc
                     cgaGTACGAA  Intron 2   CAGtctactaatgctactcaaactaa
                     tag<0-----[661  :  801]-0>atgaatgtttacatttggaacta


CAEEL-PGP-1____  960 NFCEKLDIPHKEAIKEAFIQGLSYG                     984
                      FC+ LD PH++A++E+FIQG++YG
                     KFCDYLDAPHRDALRESFIQGVAYG
H_S09_0044638__  871 attgtcggcccggccgttacgggtg                     944
```

```
                              atgaataccagactgacttagtcag
                              atcttgtcatccaacaaccaaatt


H_S09_0044638__  945 GTTTGTT  Intron                               1009
                      <2-----[945  :     ]-2>
//

Gene 1
EXONS 377 944
  Exon 377 500 phase 0
  Exon 581 660 phase 1
  Exon 802 944 phase 0
//

Making a G in phase 1 intron


>H_S09_0044638___HAECO_S09_Supercontig_0044638
_377-944bp_AA
_CAEEL-PGP-1_869-984AA

AIDFRLSTVIMTLISMLAGIVLAFFYGWQMAFLVVGILPLLGIGQALRVRVMGGKHRKNA
KDFEDSGKVAMEAIEHVRTVQALTKEEAFHQKFCDYLDAPHRDALRESFIQGVAYG
//

>H_S09_0044638___HAECO_S09_Supercontig_0044638

gcaatcgattttcgtttgagtacggtcatcatgacattgatttcaatgttagctggtatt
gtattggcattcttctatggttggcaaatggcatttttggttgttggaattcttccttta
cttggaatcggacaagctttacgagttcgtgttatgggtggcaagcatcgaaagaatgca
aaagactttgaagattctggaaaggtagctatggaagcaattgagcatgttcgtacagtc
caagctcttactaaggaggaagcattccatcaaaaattttgcgattatctggatgcccca
catcgcgacgcactacgcgaatcattcatccaaggagtagcttatgg
//
```

A.9 *HAECO*_S09_SUPERCONTIG_0046186 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1___CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence      H_S09_0046186___HAECO_S09_Supercontig_0046186

H_S09_0046186__-1406           Intron     CAG                          918
                                <2-----[     :  918]-2>


CAEEL-PGP-1____  985  CASSVLYLLNTCAYRMGLALIITDPPTMQPMRVL
                       A+S++Y+LN C+YR+GL LI++    M P RVL
                      FATSIVYVLNCCSYRLGLYLIVSS--IMMPTRVL
H_S09_0046186__ -917  tgatagtgtattttctgctcagaa  aaacaagt
                      tccctattaggcagtgtatttgg  tttccgtt
                     attatagtagcccatggtctgtcct  tggagatg


CAEEL-PGP-1____ 1019                R          VMYAITISTSTLGFATSYFPEYA
                                    R          VMYAITIS+STLGFA++YFPEY
                                    R          VMYAITISSSTLGFASAYFPEYM
H_S09_0046186__ -820 a           [aqa]         gatgaaatttacgtgtgttcgta
                     gGTGAGTT  Intron 1   TAG ttactctcccctgtcccatcaat
                      <2-----[818  :  763]-2>aagcccatgaattgccgctctgtg


CAEEL-PGP-1____ 1043 KATFAGGIIFGMLRKISKIDSLSLAGEKK
                     KA FAGGIIF ML++ S ID+L+  G+K+
                     KAAFAGGIIFNMLKQKSSIDNLTHDGKKE
H_S09_0046186__ -692 aggtgggaataacacattagacacggaag
                     acctcggtttattaaacctaatcaagaaaGTTTGAT  Intron 2
                     aatctctctccgtggaccctcccctcaag<0-----[605  :  536]


CAEEL-PGP-1____ 1072    KLYGKVIFKNVRFAYPERPEIEILK
                        KL G + FKNV+F+YPERP+IE+LK
                        KLSGAITFKNVKFSYPERPQIEVLK
H_S09_0046186__ -538    acaggaataagatatcgaccaggta
                     CAGatggctctaatatgacagcatattaGTATGTG  Intron 3
                     -0>aacttcgcatgacccgggcaaaagg<0-----[460  :  352]


CAEEL-PGP-1____ 1097    GLSFSVEPGQTLALVGPSGCGKSTVVALLERFYDTLGGEI
                        GLSF+ +PG+TLALVGPSGCGKSTVV+L+ERFYD   G++
                        GLSFTAKPGETLALVGPSGCGKSTVVSLIERFYDVKAGQV
H_S09_0046186__ -354    gtttagacggatgcggctgtgaaaggtcagcttggaggcg
                     CAGgtctccacgactcttgccgggagcttcttagtaatacgat
                     -0>tgaccgatcaggactcatttcacatcggtgatctcgtaag


CAEEL-PGP-1____ 1137                        FIDGSEIKTLNPEHTRSQIAIVSQEP
                                            ++D  + ++LNP HTRSQIAIVSQEP
                                            LLDSHDPRSLNPYHTRSQIAIVSQEP
H_S09_0046186__ -231                        ccgtcgcctcactcactcagagtcgc
                     GTAGGCG  Intron 4   CAGttacaacgctacaacgcatcttcaac
                                          <0-----[231  :  178]-0>gctaccttagtatcaaaacaccaagg


CAEEL-PGP-1____ 1163 TLFDCSIAENIIYGLD                                1178
                      LFDCSIA+NI YG++
                     ILFDCSIADNIAYGME
H_S09_0046186__  -99 actgttaggaagtgagg                               51
                     tttagctcaatcagta
                     tacccttccttacaga


H_S09_0046186__  -50 GTGATTA  Intron                                    1
                      <1-----[50   :    1]-1>
//

Gene 1
EXONS 917 51
  Exon 917 819 phase 0
  Exon 762 606 phase 2
  Exon 535 461 phase 0
  Exon 351 232 phase 0
  Exon 177 51 phase 0
//

Making a R in phase 2 intron


>H_S09_0046186___HAECO_S09_Supercontig_0046186
_917-51bp_AA
_CAEEL-PGP-1_985-1178AA

F
ATSIVYVLNCCSYRLGLYLIVSSIMMPTRVLRVMYAITISSSTLGFASAYFPEYMKAAFA
GGIIFNMLKQKSSIDNLTHDGKKEKLSGAITFKNVKFSYPERPQIEVLKGLSFTAKPGET
LALVGPSGCGKSTVVSLIERFYDVKAGQVLLDSHDPRSLNPYHTRSQIAIVSQEPILFDC
SIADNIAYGME
//

>H_S09_0046186___HAECO_S09_Supercontig_0046186

attt
gctacatctatagtgtatgtattgaactgctgctcatatcggttgggtctctatctgatt
gtcagcagtattatgatgccaacgagagttttgagagtaatgtacgccatcacaatttcg
tcatcaactcttgggttcgcctcggcctatttccctgagtatatgaaagcagctttcgct
ggcggtatcattttcaacatgcttaagcagaaatcctccatcgataacctcacccacgat
ggcaaaaaagagaaactaagcggtgctatcacgttcaaaaatgtgaaattcagctacccg
gagaggccccaaatagaagtattgaagggtttgtcattcaccgcgaaacctggcgaaacg
ttggcactcgttggcccatctggttgtggcaaaagcacagttgtctcgctgattgagcga
ttttacgatgtcaaggctggacaagtgctgctcgattcacacgaccctcgttcactgaat
ccatatcacacacgatcacaaatcgcaatcgtctcacaagagccgattctattcgactgc
tctattgccgacaatattgcatacggaatggaa
g
//
```

A.10 *HAECO*_S09_SUPERCONTIG_0047549 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence      H_S09_0047549___HAECO_S09_Supercontig_0047549

CAEEL-PGP-1____ 1162 PTLFDCSIAENIIYGLDPSSVTMAQVEEAARLANIHNFIAELPE
                     P LFD SI +NI+YGL P SV+ A+V + A+ ANIH F+ ELPE
                     PILFDRSIRDNILYGLPPGSVSEAEVHDVAQRANIHKFVMELPE
H_S09_0047549__    1 cactgatacgaactgcccgtgaggggcgggccgaacatgagtcg
                     ctttagctgaattagtccgctgacataatcagcataatttatca
                     gcaccaccgcccctcgaatcatatagttcaattctcatagagcg


CAEEL-PGP-1____ 1206                         GFETRVGDRGTQLSGGQKQRIAIARA
                                             G+ TR G++G QLSGGQKQRIAIARA
                                             GYNTRAGEKGVQLSGGQKQRIAIARA
H_S09_0047549__  133                         gtaacgggaggcttggcaccagagcg
                     GTGTGTT  Intron 1   TAGgaacgcgaagtatcggaaagtctcgc
                     <0-----[133  :  196]-0>cccatgaaaccggttggaaacccatt


CAEEL-PGP-1____ 1232 LVRNPKILLLDEATSALDTESEK
                     L+RNPKILLLDEATSALDTESEK
                     LIRNPKILLLDEATSALDTESEK
H_S09_0047549__  275 caaacaatccgggaagtgagaga
                     ttgacattttaaccgctacagaaGTGCTGT  Intron 2   CAG
                     ccatgatagtcgtcttgtcacag<0-----[344  :  411]-0>


CAEEL-PGP-1____ 1255 VVQEALDRAREGRTCIVIAHRLNTVMNADCIAVVSNGTIIEK
                     VVQEALD+A EGRTCIV+AHRL+TV+NA+CI VV  G I+EK
                     VVQEALDKASEGRTCIVVAHRLSTVVNANCIMVVQGGKIVEK
H_S09_0047549__  412 ggcggcgagtggcatagggccctaggagataaggcggaaggag
                     ttaactaaccaggcgtttcagtccttacagttttaggattaa
                     tgagcccgaaaccctctgatgggtcctctccgacaaaattaa


CAEEL-PGP-1____ 1297                 G          THTQLMSEKGAYYKLTQKQMTEK
                                     G          TH +LM  KG Y++LTQKQ T K
                                     G          THNELMQAKGVYWELTQKQTTAK
H_S09_0047549__  539               [gga]         acagcacgaggttgcacacaaga
                      GTACAGC  Intron 3   CAGgcaaattacagtagatcaaaccca
                      <1-----[539  :  599]-1>aaccaagaaatgtggcaaggggaa


CAEEL-PGP-1____ 1321 K                                              1321
                     +
                     E
H_S09_0047549__  671 g                                               673
                     a
```

```
                     a


CAEEL-PGP-1____      *
                     *
                     *
H_S09_0047549__  674 t                                               676
                     a
                     a


H_S09_0047549__  677 [677  : 1569]                                  1569
//

Gene 1
EXONS 1 673
  Exon 1 132 phase 0
  Exon 197 343 phase 0
  Exon 412 538 phase 2
  Exon 600 673 phase 1
//

Making a G in phase 1 intron


>H_S09_0047549___HAECO_S09_Supercontig_0047549
_1-673bp_AA
_CAEEL-PGP-1_1162-1321aa_End

PILFDRSIRDNILYGLPPGSVSEAEVHDVAQRANIHKFVMELPEGYNTRAGEKGVQLSGG
QKQRIAIARALIRNPKILLLDEATSALDTESEK
VVQEALDKASEGRTCIVVAHRLS
TVVNANCIMVVQGGKIVEKGTHNELMQAKGVYWELTQKQTTAKE
//

>H_S09_0047549___HAECO_S09_Supercontig_0047549

CCGATCCTATTCGACAGATCCATCCGGGACAACATCCTCTATGGCCTGCCACCAGGTTCC
GTAAGTGAAGCTGAAGTGCATGATGTCCGCACAACGTGCTAACATTCACAAATTTGTAATG
GAATTGCCCGAGGGCTACAACACACGTGCGGGAGAAAAAGGCGTCCAGTTGTCTGGTGGG
CAGAAACAACGAATCGCCATCGCACGTGCTCTCATCAGAAATCCGAAAATTTTACTGCTT
GACGAGGCTACCAGTGCTTTGGATACCGAAAGCGAAAAGGTGCTTGTTCAGGTTGTGCAA
GAGGCCCTCGACAAGGCATCAGAAGGCCGCACCTGTATCGTTGTGGCACATCGGCTGTCG
ACTGTCGTCAATGCCAATTGCATCATGGTAGTCCAAGGAGGAAAAATTGTTGAAAAAGGA
ACACACAACGAACTAATGCAAGCAAAAGGTGTGTATTGGGAGCTCACACAAAAGCAGACG
ACGGCAAAAGAA
//
```

A.11 *HAECO*_S09_SUPERCONTIG_0024971 ≈ *CAEEL*-PGP-1

```
genewise
Query protein:       CAEEL-PGP-1____CAEEL-PGP-1 K08E7 9 1-1321AA
Target Sequence      H_S09_0024971___HAECO_S09_Supercontig_0024971

CAEEL-PGP-1____ 1266 GRTCIVIAHRLNTVMNADCIAVVSNGTIIEK
                     GRTCIVIAHR +TV+NADCIAVV  G IIE+
                     GRTCIVIAHRFSTVVNADCIAVVKGGVIIEQ
H_S09_0024971__    1 gcatagagcctaaggaggtagggagggaagc
                     ggcgtttcagtgcttacagtcttaggtttaa
                     atacatttttcctcctttttttcaactgctga


CAEEL-PGP-1____ 1297                  G          THTQLMSEKGAYYKLTQKQ
                                      G          THT+LM+++G YY+LTQKQ
                                      G          THTELMAKRGFYYELTQKQ
H_S09_0024971__   94 g           [gga]          acagcagacgtttgcacac
                      GTTCGAT  Intron 1   CAGgcacattcaggtaaatcaaa
                      <1-----[95    : 1037]-1>attaatgaaatcccaccaga


CAEEL-PGP-1____ 1317 MTEKK--                                      1321

                     TVKSASE
H_S09_0024971__ 1097 agatgag                                      1117
                     ctaccga
                     ttgtcta


CAEEL-PGP-1____      *
                     *
                     *
H_S09_0024971__ 1118 t                                            1120
```

```
                     a
                     a


H_S09_0024971__            [1121 :      ]                        5666
                     2997 n [1170 : 4166]
//

Gene 1
EXONS 1 1117
  Exon 1 94 phase 0
  Exon 1038 1117 phase 1
//

Making a G in phase 1 intron


>H_S09_0024971___HAECO_S09_Supercontig_0024971
 1-1117bp_AA
_CAEEL-PGP-1_1266-1321AA_End

GRTCIVIAHRFSTVVNADCIAVVKGGVIIEQGTHTELMAKRGFYYELTQKQ
TVKSASE
//

>H_S09_0024971___HAECO_S09_Supercontig_0024971

ggacgtacatgcatagttattgctcatcgtttcagcactgtcgtcaatgctgattgtatt
gctgtcgtaaaaggcggtgtgatcattgagcaaggaactcatacagaacttatggcaaaa
cgaggtttctactacgaactcacccaaaagcaa
actgttaagtctgccagtgaa
//
```

A.12 *HAECO*_S09_SUPERCONTIG_0049649 ≈ *CAEEL*-PGP-2

```
genewise
Query protein:       CAEEL-PGP-2____CAEEL-PGP-2 C34G6 4 1-1272AA
Target Sequence      H_S09_0049649___HAECO_S09_Supercontig_0049649

H_S09_0049649__     1              Intron CAG
                    <0-----[     : 3938]-0>
                    1631 n [1091 : 2721]


CAEEL-PGP-2____   394 NMKGDISFKDVHFRYPSRKDIHVLKGISLELKAGDKIALVGSSGCGKSTI
                        + GD+S  ++ F YPS+  + V  G+S+    G  IALVG SG GKST+
                    VISGDLSVDNIIFSYPSHPKFRVANGLSFNACQGQSIALVGPSGSGKSTV
H_S09_0049649__  3939 gatggctggaaatttcaccatcggagcttagtcgctagtggctgtgaaag
                    ttcgatctaatttcacgacatgtcagtctacgagactcttgccgcgagct
                    gtcttgtgcttctctattagctcttctcctttacgtcagttatatgacaa


CAEEL-PGP-2____   444 VNLLQRFYDPTKGRV                          LIDGVDLREVN
                    VNLLQRFYD  +G +                            +DG ++  +N
                    VNLLQRFYDVQEGSI                           SVDGRNILTMN
H_S09_0049649__  4089 gacccattggcggta                          agggaaataaa
                    tattagtaataagctGTATGTA  Intron 1   CAGgtaggattcta
                    cttaaattccaatat<0-----[4134 : 4187]-0>tttagttacgt


CAEEL-PGP-2____   470 VHSLREQIGIVSQEPVLFD              G          TIYE
                       LR+ I +V QEPVLF+               G          TI E
                    AQHLRDNIALVEQEPVLFE              G          TILE
H_S09_0049649__  4221 gccccgaagtggcgcgctgg           [ggg]        aacg
                    caatgaatcttaaacttta GTAAGTT  Intron 2   CAGgctta
                    aatactcagatagacgtca <1-----[4279 : 5255]-1>ggtta


CAEEL-PGP-2____   494 NIKMGNEHATHDQVVEACKMANANDFIKRLPD              525
                    N+ +G +  T D  ++AC++ANA+ FI+RLP+
                    NVLLGTDKYTEDDAIKACRLANASQFIERLPE
```

```
H_S09_0049649__ 5270 agccgagataggggaagtccgagtctagaccg              5365
                     atttgcaaacaaactacggtcaccattagtca
                     cggatttacggtttcatccgctggaccagaag
//

Gene 1
EXONS 3938 5365
  Exon 3939 4133 phase 0
  Exon 4188 4278 phase 0
  Exon 5256 5365 phase 1
//

Making a G in phase 1 intron


>H_S09_0049649__HAECO_S09_Supercontig_0049649
_3939-5365bp_AA
_CAEEL-PGP-2_394-525AA

V
ISGDLSVDNIIFSYPSHPKFRVANGLSFNACQGQSIALVGPSGSGKSTVVNLLQRFYDVQ
EGSISVDGRNILTMNAQHLRDNIALVEQEPVLFEGTILENVLLGTDKYTEDDAIKACRLA
NASQFIERLPE
//

>H_S09_0049649___HAECO_S09_Supercontig_0049649

gtg
atttccggtgatctgtctgtggacaatattatctttttcctatccaagtcatccaaagttc
cgtgtcgctaatggcctttccttcaatgcttgtcaaggccagtctatcgcattggttggt
ccatctggatctgggaaaagcacagtagtcaatcttctacaaagattttatgacgtccaa
gaaggttcaattagtgttgatggaaggaatattttaaccatgaatgcacaacatctacgc
gataacatagcgttagttgaacaggaacccgtgcttttcgaagggacgattcttgaaaac
gtgctgctaggtactgataaatacacggaggatgatgctatcaaagcttgccgcctggcc
aatgcgtcgcaattcatcgaaaggctaccagag
//
```

A.13 *HAECO*_S09_SUPERCONTIG_0007774 = *HAECO*-PGP-2.1.A

```
genewise
Query protein:       HAECO_AAC38987_HAECO-PGP-2_1_A_AAC38987_1-1275AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0007774___HAECO_S09_Supercontig_0007774
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623

H_S09_0007774__-11145          Intron      TTCTTTTCAG
                     <0-----[     : 9871]--------0>

HAECO_AAC38987_   40 LSLATTLDYVLLAAGTLAPCVHGAGFSVLGIVLGGMTTVFLRAQNSEFV
                       LSLATTLDYVLL  GTLA CVHGAGFSVLGIVLGGMTTVFLRAQNSEFV
                       LSLATTLDYVLLVVGTLASCVHGAGFSVLGIVLGGMTTVFLRAQNSEFV
H_S09_0007774__-9870 cttgaatgtgccgggacgttgcgggttgcgagcggaaagtcagcatgtg
                     tctccctaatttttgctccgtagcgtcttgtttggtcctttgcaacatt
                     ccactagttgttgttggggtttctacaacttacttggactcatgcaact

HAECO_AAC38987_   89 LGTVSRDPEGLPALT              K          EEFDTLVR
                       LGTVSRDPEGLPALT              K          EEFDTLVR
                       LGTVSRDPEGLPALT              K          EEFDTLVR
H_S09_0007774__-9723 cgagacgcggccgcaa           [aag]       ggtgacgc
                     tgctggacagtcctcaGTGAGTT  Intron 1   CAG aatacttg
                     gctttgttagaattt <2-----[9676 : 9530]-2>gaatcaaat

HAECO_AAC38987_  113 RYCLYYLGLGFAMFATSYIQ                    IVCWET
                       RYCLYYLGLGFAMFATSYIQ                    IVCWET
                       RYCLYYLGLGFAMFATSYIQ                    IVCWET
H_S09_0007774__-9504 atttttcgtgtgatgattac                    agttga
                     gagtaatgtgtcttcccataGTAACTG  Intron 2   CAGttggac
                     gtcacctaacttgtaattag<0-----[9444 : 9165]-0>tgtggg

HAECO_AAC38987_  139 FAERITHKLRKIYLKAILRQQISWFDIQQTGNLTARLT
                       FAERITHKLRKIYLKAILRQQISWFDIQQTGNLTARLT
                       FAERITHKLRKIYLKAILRQQISWFDIQQTGNLTARLT
H_S09_0007774__-9146 tggcaacatcaatcagacccccatttgaccagacagcca
                     tcagtcaatgatatacttgaatcgtataacgatccgtc
                     ccaatctaaaattaacatgggcagtctaaaatcattac

HAECO_AAC38987_  177              D          DLERVREGLGDKLSLFIQMVSAF
                                 D          DLERVREGLGDKLSLFIQM+SAF
                                 D          DLERVREGLGDKLSLFIQMMSAF
H_S09_0007774__-9032 g          [gat]       gcgcgcggcggactctacaatgt
                     aGTAAGAA  Intron 3   TAG atagtgagtgaatctttattcct
                      <2-----[9030 : 8975]-2>ttcatttaatttaggttaaggttt

HAECO_AAC38987_  201 VAGFCVGFAY              S          WSMTLVMMVVAPF
                       VAGFCVGFAY              S          WSMTLVMMVVAPF
                       VAGFCVGFAY              S          WSMTLVMMVVAPF
H_S09_0007774__-8904 gggttggtgta           [agc]        ttaacgaagggct
                     tcgtgtgtcagGTAGGTT  Intron 4   CAG gctcttttttcct
                     gttctaacgt <2-----[8872 : 8818]-2>cgaggcgggcgggt

HAECO_AAC38987_  225 IVISANWMSKIVAT              R          TQVEQETYA
                       IVISANWMSKIVAT              R          TQVEQETYA
                       IVISANWMSKIVAT              R          TQVEQETYA
H_S09_0007774__-8777 agatgatataaggaa           [agg]        acggcgatg
                     tttccagtcatttccgGTGAGCG  Intron 5   CAG cataaacac
                     atttttggaacttt <2-----[8733 : 8614]-2>gcatagacct

HAECO_AAC38987_  249 VAGAIAEETFSSIRTVHSICGHKRELT              R
                       VAGAIAEETFSSIRTVHS+CGHKRELT              R
                       VAGAIAEETFSSIRTVHSLCGHKRELT              R
H_S09_0007774__-8585 ggggagggatttacagctttgcaagcaa             [aga]
                     tcgctcaactcctgctactggaagatcgGTTGTTT  Intron 6
                     tcttagggtcagaacacgatctaagaa <2-----[8502 : 8446]

HAECO_AAC38987_  276     FEAALEKGRQTGLVKYFYMGVGVGFGQMCTYVSYALAF
                         FEAALEKGRQTGLVKYFYMGVGVGFGQMCTYVSYALAF
                         FEAALEKGRQTGLVKYFYMGVGVGFGQMCTYVSYALAF
H_S09_0007774__-8448     tgggtgagccagcgatttaggggggtgcatatgttgtgt
AAG tacctaaggacgttaatatgtgtgtgatgcatcactct
-2>atgagggaatgactcatctgtttgattggtctgcccgtt

HAECO_AAC38987_  315              W          YGSVLIINDPALDRGRIFT
                                 W          YGSVLIINDPALDRGRIFT
                                 W          YGSVLIINDPALDRGRIFT
H_S09_0007774__-8330 t          [tgg]       tgagcaaagcgtgcgcata
                     gGTATGTG  Intron 7   CAG aggttttaacctagggttc
                      <2-----[8328 : 8262]-2>gtctagccccctagttcatca

HAECO_AAC38987_  335                       VFFAVMSGSAALGTCLPHLNTISIAR
                                           VFFAVMSGSAALGTCLPHLNTISIAR
                                           VFFAVMSGSAALGTCLPHLNTISIAR
H_S09_0007774__-8203                       gttggatgtggcgatccccaaatagc
                     GTCAGTT  Intron 8   CAGtttcttcgccctgcgtcatactctcg
                     <0-----[8203 : 8142]-0>ctttggccaatccatgattccaccta

HAECO_AAC38987_  361 GAVRSVLSVINS                          RPKIDPYSLDGIVL
                       GAVRSVLSVINS                          RPKIDPYSLDGIVL
                       GAVRSVLSVINS                          RPKIDPYSLDGIVL
H_S09_0007774__-8063 gggcagctgaaa                          ccaagctttggagc
                     gctggttcttagGTGAGTG  Intron 9   TAGgcatacactagttt
                     agaatagagttt<0-----[8027 : 7956]-0>taactctgatctgc

HAECO_AAC38987_  387 NNMRGSIRFKNVHFSYPSRRTLQ                      ILK
                       NNMRGSIRFKN HFSYPSRRTLQ                      ILK
                       NNMRGSIRFKNEHFSYPSRRTLQ                      ILK
H_S09_0007774__-7913 aaaagtactaagctttctcaatc                      ata
                     aatggctgtaaaatcaccggctaGTAGGGA  Intron 10  CAGtta
                     ctgaattccgcgcccttcaaagg<0-----[7844 : 7781]-0>aga

HAECO_AAC38987_  413 GVSLQVSAGQKIALVGSSGCGKSTNVNLLLRFYDPTRGK
                       GVSLQVSAGQKIALVGSSGCGKST VNLLLRFYDPTRGK
                       GVSLQVSAGQKIALVGSSGCGKSTIVNLLLRFYDPTRGK
H_S09_0007774__-7771 ggtccgtggcaagtggtagtgataagatttattgcaaga
                     gtctatccgaatcttgcggggaccttatttgtaaccgga
                     tgagaggtcaattggtacttagagcctaagatttgagag

H_S09_0007774__-7654 GTGAGTG  Intron 11  AAG
                     <0-----[7654 : 7536]-0>

HAECO_AAC38987_  452 VTIDDIDVCDLNVQKLREQIGVVSQ                    476
                       VTIDDIDVCDLNVQKLREQIGVVSQ
                       VTIDDIDVCDLNVQKLREQIGVVSQ
H_S09_0007774__-7535 gaaggaggtgcagcaccgcagggac                   7461
                     tctaatatgatataatgaatgttga
                     acattttgttccgaattaacttttg

H_S09_0007774__-7460 GTACGTG  Intron 12
                     <0-----[7460 :     ]-0>

Alternative splice

H_S09_0007774__-          Intron 12  CAG
                     <0-----[7460 : 6579]-0>

HAECO_AAC38987_   477 EPVLFDGTLFENIKMGYEQATMEEVQEACRV                  507
                       EPVLFDGTLFENIKMGYEQATMEEVQEACRV
                       EPVLFDGTLFENIKMGYEQATMEEVQEACRV
H_S09_0007774__-6578 gcgctggactgaaaagtgcgaagggcggtcg                 6486
                     actttagcttaatatgaaacctaataacggt
                     aagtctcaacatcggttagcagggcaagctg

H_S09_0007774__-6485 GTCAGTA  Intron 13
                     <0-----[6485 : 5117]-0>
                       10 n [6403 : 6394]

Alternative splice

H_S09_0007774__-          Intron 12  CAG
                     <0-----[7460 : 5511]-0>
                       10 n [6403 : 6394]

HAECO_AAC38987_   477 EPVLFDGTLFENIKMGYEQATMEEVQEACRV                  507
                       EPVLFDGTLFENIKMGYEQATMEEVQEACRV
                       EPVLFDGTLFENIKMGYEQATMEEVQEACRV
H_S09_0007774__-5510 gcgctggattgaaaagtgcgaagggcggtcg                 5418
                     actttagcttaatatgaaacctaataacggt
                     aagtctcaacatcggttagcagggcaatctg

H_S09_0007774__-5417 GTCAGTA  Intron 13
                     <0-----[5417 : 5117]-0>

H_S09_0007774__-          Intron 13  TAG
                     <0-----[     : 5117]-0>

HAECO_AAC38987_   508 ANAADFTKRLPEGYGTRVGER
                       ANAADF KRLPEGYGTRVGER
                       ANAADFIKRLPEGYGTRVGER
H_S09_0007774__-5116 gagggtaacccggtgacgggc
                     caccattagtcagagcgtgag
                     gttccccaataatcccattat

HAECO_AAC38987_   529 GVQLSGGQKQRI              A          IARAIIKNPRI
                       GVQLSGGQKQRI              A          IARAIIKNPRI
                       GVQLSGGQKQRI              A          IARAIIKNPRI
H_S09_0007774__-5053 ggctaggcaccag           [gcc]        agcgaaaacca
                     gtatgggaaagt GTAAGTA  Intron 14  CAGctcgcttaacgt
                     tggatcaaggat <1-----[5016 : 4869]-1>cattgccgctca

HAECO_AAC38987_   553 LLLDEATSALDTEAESIVQEALEK                        AQ
                       LLLDEATSALDTEAESIVQEALEK                        AQ
                       LLLDEATSALDTEAESIVQEALEK                        AQ
H_S09_0007774__-4833 cccgggaagcgagggtagccggcga                        gc
                     tttaaccgctacacacttaactaaGTATGCT  Intron 15  CAGca
                     ggctaccttacaagaacgagtggg<0-----[4761 : 4677]-0>ta

HAECO_AAC38987_   579 KGRTTVIVAHRLSTIRNVDQIFVFK                       N
                       KGRTTVIVAHRLSTIRNVDQIFVFK                       N
                       KGRTTVIVAHRLSTIRNVDQIFVFK                       N
H_S09_0007774__-4670 agaaagaggccctaaaaggcatgta                        a
                     aggcctttcagtcctgataattttaGTTAATA  Intron 16  CAGa
                     agaacctagttgttcacgtgtctcg<0-----[4595 : 4447]-0>c

HAECO_AAC38987_   605 GTIVEQGTHAELMNKRGVFFEMTQAQVLRQEKEEEVL
                       GTIVEQGTHAELMNKRGVFFEMTQAQVLRQEKEEEVL
                       GTIVEQGTHAELMNKRGVFFEMTQAQVLRQEKEEEVL
H_S09_0007774__-4443 gaaggcgacggtaacggttgaacgcgcccgagggt
                     gcttaagcaattaagtttatcacattgaaaaatt
                     agctggcttcgggcataactagtaaaccaaggagata

HAECO_AAC38987_   642              D          SDAESDVVSPDIALPHLSSLRSR
                                 D          SDAESDVVSPDIALPHLSSLRSR
                                 D          SDAESDVVSPDIALPHLSSLRSR
H_S09_0007774__-4332 g          [gat]       agggtgggtcgagtcccatcctc
                      GTCAGAC  Intron 17  CAGagacacattccatctcatgctgcg
                      <1-----[4331 : 3361]-1>tctgactcgagttaactttatact

HAECO_AAC38987_   666 KESTRSAISAVPSVRSMQIEME                        DLRA
                       KESTRSAISAVPSVRSMQIEME                        DLRA
                       KESTRSAISAVPSVRSMQIEME                        DLRA
H_S09_0007774__-3289 agtaaagatggcagcaacagag                          gccg
                     aaccggctcctcgtggtatataGTGAGTT  Intron 18  TAGatgc
                     aacaattccgccctatgacagg<0-----[3223 : 2703]-0>cttc

HAECO_AAC38987_   692 KPTPMSKIFYFNRDKWGYFILGLIACIITGTVTPTFAVLYAQIIQ
                       KPTPMSKIFYFNRDKWGYFILGLIACIITGTVTPTFAVLYAQIIQ
                       KPTPMSKIFYFNRDKWGYFILGLIACIITGTVTPTFAVLYAQIIQ
H_S09_0007774__-2690 acacataatttacgatgttatgcagtaaagagacatggttgcaac
                     accctcattatagaaggatttgttcgttcgctccctcttacatta
                     aataggatcttctcagatctgacccttttattagatatatggcag

HAECO_AAC38987_   737                       VYSEPVDQMKGHVLFWCGAFIVIGLV
                                            VYSEPVDQMKGHVLFWCGAFIVIGLV
                                            VYSEPVDQMKGHVLFWCGAFIVIGLV
H_S09_0007774__-2555                        gttgcggcaagcgctttggtagagcg
                     GTACGTT  Intron 19  CAGtacactaatagatttgggcttttgtt
                     <0-----[2555 : 2191]-0>acgatttagactggcgtatccccttca
                       10 n [2419 : 2410]

HAECO_AAC38987_   763 HAFAFFFS                          AICLGRCGEALTKKLRFE
                       HAFAFFFS                          AICLGRCGEALTKKLRFE
                       HAFAFFFS                          AICLGRCGEALTKKLRFE
H_S09_0007774__-2112 cgtgtttt                           gattgctgggtaaatctg
                     actcttcGTCAGTT  Intron 20  CAGctgtggggactcaatgta
                     cttgctcg<0-----[2088 : 1891]-0>tttgatccagagaaatcg

HAECO_AAC38987_   789 AFKNLLRQNVGFYDDIRHGTGKLCTRFATDAPNVRY
                       AFKNLLRQ+VGFYDDIRHGTGKLCTRFATDAPNVRY
                       AFKNLLRQDVGFYDDIRHGTGKLCTRFATDAPNVRY
H_S09_0007774__-1836 gtaaccccgggttggaccgagactactgaggcagct
                     ctaattgaatgtaaatgagcgatgcgtccaccatga
                     gcgctgagtgaccctcactctactgattatactcat

HAECO_AAC38987_   825                       VFTRLPGVLSSVVTIIGALVIGFIFG
                                            VFTRLPGVLSSVVTIIGALVIGFIFG
                                            VFTRLPGVLSSVVTIIGALVIGFIFG
H_S09_0007774__-1728                        gtacccggcttggaaaggtgagtatg
                     GTGAGTG  Intron 21  GAGttcgtcgttccttctttgctttgtttg
                     <0-----[1728 :  364]-0>gctatgtgtagggcatatgttacccg
```

```
HAECO_AAC38987_  851 WQLALILMVMVPLIIGSGYFEMRMQ                      875
                     WQLALILMVMVPLIIGSGYFEMRMQ
                     WQLALILMVMVPLIIGSGYFEMRMQ
H_S09_0007774__ -285 tccgtacagagctaagagttgacac                     211
                     gatcttttttttctttgggatatgta
                     gggtgttgggaggccttaccggcgg


H_S09_0007774__ -210 GTTGGTG  Intron                                1
                     <0-----[210  :     ]-0>
//

Gene 1
EXONS 9870 211
  Exon 9870 9677 phase 0
  Exon 9529 9445 phase 2
  Exon 9164 9031 phase 0
  Exon 8974 8873 phase 2
  Exon 8817 8734 phase 2
  Exon 8613 8503 phase 2
  Exon 8445 8329 phase 2
  Exon 8261 8204 phase 2
  Exon 8141 8028 phase 0
  Exon 7955 7845 phase 0
  Exon 7780 7655 phase 0
  Exon 7535 7461 phase 0

  Exon 6578 6486 phase 0

  Exon 5510 5418 phase 0

  Exon 5116 5017 phase 0
  Exon 4868 4762 phase 1
  Exon 4676 4596 phase 0
  Exon 4446 4332 phase 0
  Exon 3360 3224 phase 1
  Exon 2702 2556 phase 0
  Exon 2190 2089 phase 0
  Exon 1890 1729 phase 0
  Exon  363  211 phase 0
//

Making a K in phase 2 intron
Making a D in phase 2 intron
Making a S in phase 2 intron
Making a R in phase 2 intron
Making a R in phase 2 intron
Making a W in phase 2 intron
Making a A in phase 1 intron
Making a D in phase 1 intron


>H_S09_0007774___HAECO_S09_Supercontig_0007774
_9870-7461bp_AA
_5510-5418bp_AA_Alternative_splice_6578-6486bp_AA_Alternative_splice
_5116-211bp_AA
_HAECO-PGP-2_1_A_40-476AA_477-507AA_Alternative_splices_508-875AA

LSLATTLDYVLLVVGTLASCVHGAGFSVLGIVLGGMTTVFLRAQNSEFVLGTVSRDPEGL
PALTKEEFDTLVRRYCLYYLGLGFAMFATSYIQIVCWETFAERITHKLRKIYLKAILRQQ
ISWFDIQQTGNLTARLTDDLERVREGLGDKLSLFIQMMSAFVAGFCVGFAYSWSMTLVMM
VVAPFIVISANWMSKIVATRTQVEQETYAVAGAIAEETFSSIRTVHSLCGHKRELTRFEA
ALEKGRQTGLVKYFYMGVGVGFGQMCTYVSYALAFWYGSVLIINDPALDRGRIFTVFFAV
MSGSAALGTCLPHLNTISIARGAVRSVLSVINSRPKIDPYSLDGIVLNNMRGSIRFKNEH
FSYPSRRTLQILKGVSLQVSAGQKIALVGSSGCGKSTIVNLLLRFYDPTRGKVTIDDIDV
CDLNVQKLREQIGVVSQ

EPVLFDGTLFENIKMGYEQATMEEVQEACRV
```

```
EPVLFDGTLFENIKMGYEQATMEEVQEACRV

ANAADFIKRLPE
GYGTRVGERGVQLSGGQKQRIAIARAIIKNPRILLLDEATSALDTEAESIVQEALEKAQK
GRTTVIVAHRLSTIRNVDQIFVFKNGTIVEQGTHAELMNKRGVFFEMTQAQVLRQEKEEE
VLDSDAESDVVSPDIALPHLSSLRSRKESTRSAISAVPSVRSMQIEMEDLRAKPTPMSKI
FYFNRDKWGYFILGLIACIITGTVTPTFAVLYAQIIQVYSEPVDQMKGHVLFWCGAFIVI
GLVHAFAFFFSAICLGRCGEALTKKLRFEAFKNLLRQDVGFYDDIRHGTGKLCTRFATDA
PNVRYVFTRLPGVLSSVVTIIGALVIGFIFGWQLALILMVMVPLIIGSGYFEMRMQ
//

>H_S09_0007774___HAECO_S09_Supercontig_0007774
_9870-7461bp
_Alternative_splice_6578-6486bp_5510-5418bp
_5116-211bp

ctctccttagccactacattggattatgtgcttcttgtggttggtacgctggcgtcgtgt
gttcatggcgctggattctcagtactcggtattgtactcggtggtatgacgacagtcttt
ctcagagctcagaactcagaattcgttctgggcactgttagtcgggatcctgaagggcta
ccagctcttactaaggaagaatttgacacactagtacgtaggtattgcttatactacctt
ggattaggctttgctatgtttgcaacatcttatatacagattgtgtgttgggagacgttc
gccgaacgaattacccataaattacgaaaaatttatctaaaagccatacttcggcagcag
atctcatggtttgacattcaacaaacaggaaatctcacagctcgtctaaccgatgatctc
gaacgtgttcgtgaaggacttggtgataaactgtcgctttttatacaaatgatgtctgct
tttgtggctggtttctgtgtaggattcgcgtatagctggtcaatgacgctcgtgatgatg
gtcgtggcgccgtttatagttatttctgctaattggatgtcaaaaatcgttgctactagg
acccaagttgaacaggaaacctacgctgttgccggtgctatagcggaggagactttctca
tcgatacgaaccgtacactcgttatgtggccataaaagagagctaacaagatttgaggca
gcgttggagaaaggacgtcagacaggccttgtcaaatatttctatatgggtgttggtgtg
ggatttggtcagatgtgtacctatgtgtcctacgccttggcttttggtatggcagtgta
ctgatcatcaacgaccctgcattggatcgtggccgaattttcacagtctttttgctgtg
atgtccggctcagcagctctcggcacatgtctgccacatcttaacaccatatccatcgct
cgaggagcggtacgaagtgtactgtcagtgattaatagtcgtccaaaaatcgatccctat
tcgttagatggcattgtgctcaacaatatgagaggatctattcgcttcaagaacgagcac
ttctcctatccttcccgaagaacattgcagatattgaaaggtgtgtcactgcaagtgtcg
gctggccaaaaaattgctttggtgggttcaagcggttgtggaaagtcaacgatcgtcaat
ttattattgagattttatgatccgacaaggggaaaggtaaccatagatgatattgatgtg
tgtgatctcaacgtgcaaaaacttcgtgaacaaatcggtgttgttagtcag

gaaccagtg
cttttcgatggcacattattcgaaaatatcaagatgggttatgaacaggccacaatggag
gaggtccaagaagcttgccgtgtg

gaaccagtg
cttttcgatggcacattattcgaaaatatcaagatgggttatgaacaggccacaatggag
gaggtccaagaagcttgccgtgtg

gcgaatgctgccgacttcatcaaacgacttccagaa
ggttacggcacccgagttggtgaacgtggtgtgcagttaagtggcggacaaaagcagcga
attgccatagctcgtgcgatcatcaagaaccctcgcatactgctgctcgatgaagccacc
agtgctctagacacagaagcggaatcaatcgtgcaagaggctctggagaaggctcaaaaa
gggagaacaaccgtcattgtagcgcatcgtctgtctactatcagaaacgtggatcagatt
ttcgttttcaagaacggaacgatcgttgagcagggcactcatgccgagttgatgaacaaa
cgtggagtattctttgaaatgactcaagcacaagtcctccgacaagagaaggaagaggaa
gttttagatagcgatgcggaatccgatgtcgtgtcaccggatattgcattaccccatctt
agttcacttcgatcccgtaaagaatccacaagaagtgctatctccgcggtccccagcgtt
cgaagtatgcaaatcgaaatggaggaccttcgtgccaaaccaactccaatgtcgaaaatt
ttctattttaaccgtgacaaatgtgggatatttcatttttgggactcatcgcctgtattatt
actggaactgttacaccgacatttgcagttttatatgcgcagatcatacaggtatactcg
gaacctgttgatcaaatgaaaggccatgtgctgttctggtgtggagctttcatcgtcatt
ggtctcgtacacgctttgcgttcttttctcggctatttgtttgggacgttgcggcgaa
gcgttaacgaaaaaattacgtttcgaggcgttcaagaaccttctgcgacaggatgtggga
ttctacgacgatatccgacacggtaccggtaaactctgtacgcgatttgctacagatgca
cccaatgtccgatatgtgttcactcgacttccgggtgtgctttcatcggtggtgaccata
attggagctttggttattggattcatcttcgggtggcagctggctttgattcttatggtg
atggtaccgttgatcatcggtagtggatacttcgagatgcgcatgcag
//
```

A.14 *HAECO*_S09_SUPERCONTIG_0038725 = *HAECO*-PGP-2.1.A

```
genewise
Query protein:      HAECO_AAC38987_HAECO-PGP-2 1 A AAC38987 1-1275AA
Target Sequence     H_S09_0038725___HAECO_S09_Supercontig_0038725

H_S09_0038725__     1              Intron   CAG
371
                    <0-----[     :  371]-0>


HAECO_AAC38987_ 1025 KISGHISFRNVYFNYPTRRQIRVLRGLNLE
                     KISGHISFRNVYFNYPTRRQIRVLRGLNLE
                     KISGHISFRNVYFNYPTRRQIRVLRGLNLE
H_S09_0038725__  372 aatgcattcagttatcaaacaagccgcacg
                     atcgatctgatataaccggatgttggtata
                     actttcgcctctcttgaaagcaactatcag


HAECO_AAC38987_ 1055                         INPGTTVALVGQSGCGKSTVMALLER
                                             INPGTTVALVGQSGCGKSTVMALLER
                                             INPGTTVALVGQSGCGKSTVMALLER
H_S09_0038725__  462                         aacgaaggcggctgtgaaagagtttgc
                    GTAACTA  Intron 1   TAGtacgcctcttgacgggagcttcttag
                    <0-----[462  :  615]-0>attcggagttggtttaactgggggag


HAECO_AAC38987_ 1081 FYNQNKGVI                          TVDGENIRNMNIRNLRE
                     FYNQNKGVI                          TVDGENIRNMNIRNLRE
                     FYNQNKGVI                          TVDGENIRNMNIRNLRE
H_S09_0038725__  694 ttacaagga                          aggggaaaaaaacaccg
                     taaaaagttGTAAGTA  Intron 2   CAGctagaatgatatgatga
                     tctacgcgt<0-----[721  : 1001]-0>ggccaccacgcactttg


HAECO_AAC38987_ 1107 QVCIVSQEPTLFDCTIMENICYGLDDPKPSYEQVVAAAKMANIHNFVLG
                     QVCIVSQEPTLFDCTIMENICYGLDDPKPSYEQVVAAAKMANIHNFVLG
                     QVCIVSQEPTLFDCTIMENICYGLDDPKPSYEQVVAAAKMANIHNFVLG
H_S09_0038725__ 1053 cgtagacgcgcactgtaaagaattgcggcacttgcgggggagaagaacatgcg
                     atgttgaaccttagcttaatgagtaacaccaaattcccatcataatttg
                     agttacgaaggcctccgacctctctccggccagtttaaaggctcttgga
```

```
HAECO_AAC38987_ 1156 LPE                                          1158
                     LPE
                     LPE
H_S09_0038725__ 1200 ccg                                         1208
                     tca
                     aag


H_S09_0038725__ 1209 GTGAGTG  Intron
2643
                    <0-----[1209 :     ]-0>
//

Gene 1
EXONS 372 1208
  Exon 372 461 phase 0
  Exon 616 720 phase 0
  Exon 1002 1208 phase 0
//

>H_S09_0038725___HAECO_S09_Supercontig_0038725
_372-1208bp_AA
_HAECO-PGP-2.1.A_1025-1158AA

KISGHISFRNVYFNYPTRRQIRVLRGLNLEINPGTTVALVGQSGCGKSTVMALLERFYN
QNKGVITVDGENIRNMNIRNLREQVCIVSQEPTLFDCTIMENICYGLDDPKPSYEQVVAA
AKMANIHNFVLGLPE
//

>H_S09_0038725___HAECO_S09_Supercontig_0038725

AAAATCTCTGGTCATATCTCGTTCCGCAATGTCTATTTCAATTATCCGACAAGAAGA
CAGATCAGAGTACTCCGTGGACTTAACCTAGAGATAAATCCTGGCACGACGGTAGCGCTT
GTTGGGCAGTCTGGTTGTGGAAAAAGCACTGTGATGGCGTTGTTGGAACGGTTTTACAAT
CAAAACAAGGGCGTGATTACGGTGGACGGCGAAAACATCAGAAACATGAACATACGCAAT
CTTCGTGAGCAAGTGTGTATTGTAAGCCAGGAACCAACGCTGTTCGACTGTACCATCATG
GAAAACATCTGTTACGGTCTCGATGACCCCAAGCCGTCCTACGAACAGGTTGTTGCTGCA
GCAAAAATGGCGAACATTCACAATTTTGTGCTGGGACTACCAGAG
//
```

A.15 *HAECO*_S09_SUPERCONTIG_0046285 = *HAECO*-PGP-2.1.A

```
genewise
Query protein:       HAECO_AAC38987_HAECO-PGP-2 1 A AAC38987 1-1275AA
Target Sequence      H_S09_0046285___HAECO_S09_Supercontig_0046285

HAECO_AAC38987_ 1177 QRIAIARALIRDPPILLLDEATSALDTESEK
                     +RIAIARALIRDPPILLLDEATSALDTESEK
                     TRIAIARALIRDPPILLLDEATSALDTESEK
H_S09_0046285__-1194 aaagagagcacgccacccgggaagcgagaga
                     cgtctcgcttgacctttaaccgctacagaa
                     gaacacaggtatgtatggtggacggtcgtag


HAECO_AAC38987_ 1208                         IVQDALEVARQGRTCLVIAHRLSTIQ
                                             IVQDALEVARQGRTCLVIAHRLSTIQ
                                             IVQDALEVARQGRTCLVIAHRLSTIQ
H_S09_0046285__-1101                         agcggcgggccgaatcgagccctaac
                     GTGCGGT  Intron 1   CAGttaactatcgaggcgtttcagtccta
                     <0-----[1101 :  427]-0>cgaccagttcatagctatctcttata


HAECO_AAC38987_ 1234 DSDVIVMIQEGKATDR              G          THEHLLM
                     DSDVIVMIQEGKATDR              G          THEHLLM
                     DSDVIVMIQEGKATDR              G          THEHLLM
H_S09_0046285__ -348 gaggagaacggagagag          [ggc]        acgctca
                     agatttttaagaccag GTATGCC  Intron 2   CAGgcaaattt
                     ctccaggcgggataca <1-----[299  :  230]-1>cttatagg


HAECO_AAC38987_ 1258 KNDLYKRLCETQRLVESQ                          1275
                     KNDLYKRLCETQRLVESQ
                     KNDLYKRLCETQRLVESQ
H_S09_0046285__ -206 aagctacctgacccggtc                           153
                     aaataagtgacagttaca
                     gctacagacaaaactaaa


HAECO_AAC38987_      *
```

```
                     *
H_S09_0046285__ -152 t                                            150
                     g
                     a


H_S09_0046285__ -149 [149  :    1]                                  1
//

Gene 1
EXONS 1191 153
  Exon 1194 1102 phase 0
  Exon 426 300 phase 0
  Exon 229 153 phase 1
//

Making a G in phase 1 intron

>H_S09_0046285___HAECO_S09_Supercontig_0046285
_1194-153bp_AA
_HAECO-PGP-2.1.A_1177-1275AA_End

T
RIAIARALIRDPPILLLDEATSALDTESEKIVQDALEVARQGRTCLVIAHRLSTIQDSDV
IVMIQEGKATDRGTHEHLLMKNDLYKRLCETQRLVESQ
//

>H_S09_0046285___HAECO_S09_Supercontig_0046285
acg
agaatagccatagccagagcgctgattcgagatccgcctatacttctgctggatgaggcg
acaagcgcgctggataccgagagtgaaaagatcgtgcaagacgccctagaggttgctcgc
caaggtagaacgtgccttgtaattgcccatcgcctttctacaattcaagacagtgacgtc
atagtgatgatccaggaggggaaagctacagacagaggcactcatgaacatttactgatg
aagaacgatctatacaaacggctatgcgaaacacaacgactcgttgaatcacaa
//
```

A.16 *HAECO*_S09_SUPERCONTIG_0005977 = *HAECO*-PGP-2.1.A

```
genewise
Query protein:      HAECO_AAC38987_HAECO-PGP-2 1 A AAC38987 1-1275AA
Target Sequence     H_S09_0005977___HAECO_S09_Supercontig_0005977

H_S09_0005977_-36694          Intron     CAG
                    <0-----[      :20571]-0>
                      10 n [34519:34510]
                     663 n [23088:22426]


HAECO_AAC38987_ 1208 IVQDALEVARQGRTCLVIAHRLSTIQDSDVIVMIQEGKATDR
                     IVQDALEVARQGRTCLVIAHRLSTIQDSDVIVMIQEGKATDR
                     IVQDALEVARQGRTCLVIAHRLSTIQDSDVIVMIQEGKATDR
H_S09_0005977_-20570 agcggcgggccgaatcgagccctaacgaggagaacggagaga
                     ttaactatcgaggcgtttcagtcctaagatttttaagaccag
                     cgaccagttcatagctatctcttatactccaggcgggataca


HAECO_AAC38987_ 1250                 G          THEHLLMKNDLYKRLCETQRLVE
                                     G          THEHLLMKNDLYKRLCETQRLVE
                                     G          THEHLLMKNDLYKRLCETQRLVE
H_S09_0005977_-20444 g          [ggc]          acgctcaaagctacctgacccgg
                      GTATGCC  Intron 1   CAGgcaaatttaaataagtgacagtta
                      <1-----[20443:20374]-1>cttatagggctacagacaaaacta


HAECO_AAC38987_ 1274 SQ                                              1275
                     SQ
                     SQ
H_S09_0005977_-20302 tc                                           20297
                     ca
                     aa


HAECO_AAC38987_      *
                     *
                     *
H_S09_0005977_-20296 t                                            20294
```

```
                     g
                     a


H_S09_0005977_-20293       [20293:    1]                                1
                      151 n [20161:20011]
                       58 n [11432:11375]
                      680 n [8867 : 8188]
                      396 n [7306 : 6911]
                       10 n [4124 : 4115]
                       13 n [2094 : 2082]
//

Gene 1
EXONS 20570 20297
  Exon 20570 20444 phase 0
  Exon 20373 20297 phase 1
//

Making a G in phase 1 intron


>H_S09_0005977___HAECO_S09_Supercontig_0005977
_20570-20297bp_AA
_HAECO-PGP-2.1.A_1208-1275AA_End

IVQDALEVARQGRTCLVIAHRLSTIQDSDVIVMIQEGKATDRGTHEHLLMKNDLY
KRLCETQRLVESQ
//

>H_S09_0005977___HAECO_S09_Supercontig_0005977

atcgtgcaagacgccctagaggttgctcgccaaggtagaacgtgc
cttgtaattgcccatcgcctttctacaattcaagacagtgacgtcatagtgatgatccag
gaggggaaagctacagacagaggcactcatgaacatttactgatgaagaacgatctatac
aaacggctatgcgaaacacaacgactcgttgaatcacaa
//
```

A.17 *HAECO*_S09_SUPERCONTIG_0001706 ≈ *CAEEL*-PGP-3

```
genewise
Query protein:       CAEEL-PGP-3___CAEEL-PGP-3 ZK455 7 1-1268AA
Target Sequence      H_S09_0001706___HAECO_S09_Supercontig_0001706

H_S09_0001706__     1          Intron     CAG
                               <0-----[     :  380]-0>


CAEEL-PGP-3____   117 NSCLYTLCERRLHCIRKKYLKSVLRQDAKW
                      N+ LYTLCERR+H IR +YL++VLRQD  W
                      NASLYTLCERRIHSIRARYLRAVLRQDMTW
H_S09_0001706__   381 agactactgccactaagatccggcccgaat
                      acgtactgaggtactgcgatgcttgaatcg
                      tgtgtggcgagacccacattaaagtgcgag


CAEEL-PGP-3____   147 FDETTIGGLTQKMS              S          GIEKIKDGI
                      +D+   G LT KMS              S          G+E+IKDGI
                      LDQQQTGALTMKMS              S          GMERIKDGI
H_S09_0001706__   471 tgcccaggcaaaaaa          [agt]        gagcaagga
                      taaaacgctctatggGTCAGTT  Intron 1   TAG gtagtaagt
                      gtaaacaatcgagc <2-----[515  :  810]-2>tagaaagttc


CAEEL-PGP-3____   171 GDKVGVLVGGVATFISGVSIGFYM              C
                      GDK+G+++  + +FI G+S+GFY+
                      GDKLGLILASIGSFIGGISLGFYL             S
H_S09_0001706__   839 ggacgcacgtagataggatcgttta          [agt]
                      gaatgtttcctggttggtctgtatgGTAAGTT  Intron 2
                      acgtaatcagtatcctacactctg <2-----[913  : 1186]


CAEEL-PGP-3____   195     WQLTLVMMITVPLQLGSMYLSAK
                          W++TLVM+ITVPL +G+   S K
                          WRMTLVMLITVPLLIGATQFSGK
H_S09_0001706__  1184     tcaatgacaagctcaggacttga
                      CAG ggtctttttctctttgccatcgaGTGAGAT  Intron 3
                      -2>tgtgggagccgaaggaaagattag<0-----[1257 : 2617]


CAEEL-PGP-3____   219    HLNRATKNEMSAYSNAGGMANEVIAGIRTVMAFNAQPFEIN
                         L+RA+K E  AYS+A  +ANEVIAGIRTVMAFNAQPFEI+
                        LLSRASKMENYAYSSAAALANEVIAGIRTVMAFNAQPFEIH
H_S09_0001706__  2615    tctcgtaagatgtttgggcgaggaggacagagtagcctgacc
                      CAGttcgccataaacacccccctcaattcgtgcttctacactatag
                      -0>atgtatagaccgcaagactccagccaaagtgccctaatgtc


CAEEL-PGP-3____   260              R            YAHQLNEARRMGIRKAIIL
                                   R              Q  +ARR  +   ++
                                   R            KI-QWYDARRRSLNHGFLE
H_S09_0001706__  2743          [cgg]         aa cttggcccacacgtcgg
                      GTAAGCG  Intron 4   CAG at agaacggggtaagtta
                      <2-----[2743 : 6500]-2>gga ggttagctcgttataa
                        10 n [3382 : 3391]


CAEEL-PGP-3____   280               A          ICTAFPLMLMFTCMAVAFWYGAT
                                               +C  +      T    ++  G+
                                    V          LCCHLANCWESTECLNSLASGSL
H_S09_0001706__  6557          [gtt]         cttccgattgtagttattgtgat
                      GTACTAA  Intron 5   CAGttggatcaggaccagtactccggt
                      <1-----[6557 : 8596]-1>tcctttcttggtaatatcaaacca


CAEEL-PGP-3____   297 AFWYGAT                                          303
                      ++  G+
                      SLASGSL
H_S09_0001706__  8647 ttgtgat                                        8667
                      ctccggt
                      caaacca
                            10 n [8696 : 8705]


CAEEL-PGP-3____   304 LAAAGAVSSGAV                                    315
H_S09_0001706__       ------------


CAEEL-PGP-3____   316 FAVFWAVLIGTRRLGEAAPHLGAITGARLAIHD
                      FAVFW+V +GTRRL + AP LGA+ GA++A  D
                      FAVFWSVFLGTRRLSDVAPQLGAFLGAKIAAAD
H_S09_0001706__  8716 tggtttgttgacccagggcccggtcggaagggg
                      tcttgctttgcggtgatccatgcttgcatccca
                      ttccgaacgtaatgttgtaataacgaaatttat


CAEEL-PGP-3____   349 IFKVIDH                          EPEIKCTSSEGKIPEKIQG
                      IF VID                            PEI   S++G  PE++ G
                      IFAVIDR                         VPEIDPMSNDGLTPEEFVG
H_S09_0001706__  8815 atggaga                         gcgagcaaaggcacggtgg
                      ttcttagGTAGGGT  Intron 6   TAGtcatactgaagtccaattg
                      ttttttg<0-----[8836 : 8903]-0>gaactggccctacaggtaa


CAEEL-PGP-3____   375 KLTFDGIEFTYPTRPELKILKGVSF                       399
                      +LTF  I FTYP+RP ++IL GVSF
                      RLTFSNIHFTYPSRPTVEILDGVSF
H_S09_0001706__  8961 ccatаaactatctacaggacgggat                      9035
                      gtctgatatcaccgcctattagtgt
                      taccccttcttgtaagagttctccc


CAEEL-PGP-3____   400 E                                               400

                      !!!!
H_S09_0001706__  9036 tgag                                          9039


H_S09_0001706__  9040 GTGAGCG  Intron 7   TAG
                      <0-----[9040 : 9102]-0>


CAEEL-PGP-3____   401 VNPGETVALVGHSGCGKSTSIGLLMRFYNQCAGM
                      VNPGETVALVGHSGCGKST I LL+RFY Q AGM
                      VNPGETVALVGHSGCGKSTIISLLLRFYEQSAGM
H_S09_0001706__  9103 gacggaggtggctgtgataaaatcccttgcagga
                      tacgactcttgacgggaccttgtttgtaaagcgt
                      tcttatatgtgtgataaggcccgacaccgataag


CAEEL-PGP-3____   435                          IKLDGIPIQEYNIRWLRSTIGIVQQE
                                               + LDGIP+++YN++W RS +G+VQQE
                                               VALDGIPLRDYNVKWWRSVVGVVQQE
H_S09_0001706__  9205                          ggcggacccgtagattctgggggccg
                      GTTTGTA  Intron 8   TAGtctagtctgaaatagggcttgttaaa
                      <0-----[9205 : 9778]-0>gcttttagatctgaggtattttaaaa


CAEEL-PGP-3____   461 PIIFVATVAENI              R          MGDVLITDQDI
                      P+IF ATVAEN+              R          MGD  +TD+D+
                      PVIFCATVAENV              R          MGDDSLTDEDV
H_S09_0001706__  9857 cgattgagggagc          [cga]        agggacagggg
                      ctttgcctcaatgGTAAGTA  Intron 9   CAG tgaagtcaaat
                      agactcgctatg <2-----[9895 :10587]-2>agttcctttgct


CAEEL-PGP-3____   485 EEACKMANAHEFICKLSD                          RYDTVIGA
                      EEACK+ANA  FI KLS+                           ++TVIG
                      EEACKLANALGFINKLSE                         GFNTVIGE
H_S09_0001706__10622 gggtacgagtgtaaatag                          gtaagagg
                      aacgatcactgttaatgaGTACGTC  Intron 10  TAGgtacttga
                      aaatatactgccctagtg<0-----[10676:10734]-0>acctgcta


CAEEL-PGP-3____   511 GAVQLSGGQKQRVAIARAIVRKPQILLLDEATSALDTESERMVQTALDK
                      GAVQLSGGQKQR+AIARA+VR PQILLLDEATSALDTESER VQ ALDK
                      GAVQLSGGQKQRIAIARALVRNPQILLLDEATSALDTESERAVQEALDK
H_S09_0001706__10759 gggcctggcaccagagagcgaaccatccgggaagtgagagcggcggcga
                      gctatcggaaagtctcgcttgacatttaaccgctacagagctaactaa
                      acggaaccggaactacacgcactatgtgtagcttgcgatattaaaatcg


CAEEL-PGP-3____   560                          ASEGRTTLCIAHRLSTIRNASKILVF
                                               A E RTTLCIAHRLSTIR++ KI+VF
                                               ARENRTTLCIAHRLSTIRDSDKIIVF
H_S09_0001706__10906                           gcgaaaactagcactaaagtgaaagt
                      GTCAGAA  Intron 11  CAGcgaagcctgtcagtcctgacaatttt
                      <0-----[10906:10992]-0>ttacgttacatcgtgccgctcactac


CAEEL-PGP-3____   586 DQGLIAER              G          THDELISKDDGIYAS
                      D+G I E+              G          THDEL+S +DG+Y S
                      DEGHIVEQ              G          THDELMSIEDGVYRS
H_S09_0001706__11071 gggcaggcg          [ggc]        acggcataggggtca
                      aagattaa GTGAGCA  Intron 12  CAGgcaaattctaagtagg
                      cgatacaa <1-----[11096:11961]-1>cgttaggatgttgttc


CAEEL-PGP-3____   610 MVKAQEIERAKEDTTLD              D          EEDEKT
                      MVKAQ IE+ +EDTTLD              D           +  +
                      MVKAQAIEKGEEDTTLD              D          VDPTEI
H_S09_0001706__12009 agagcgagagggaacgg          [gac]        ggcaga
                      ttacactaagaaaccta GTAGGTG  Intron 13  CAGataccat
                      ggagagcaacaactcct <1-----[12061:12154]-1>cgtgaac


CAEEL-PGP-3____   634 HRSFHRDSVTSDEERELQQSLARDSTRLRQSMISTTTQVPEWEIE
                       R   R  V S+++ + +   AR+S RLRQSM+S +TQ PEWEIE
                      RRGVSR--VVSEDDSQSRADRARESARLRQSMVSASTQEPEWEIE
H_S09_0001706__12175 ccggtc  ggtgggacacggcgagtgacccaagagtacgcgtgag
                      gggtcg  ttcaaagaggcagcgaccgtgagttgcccaacagata
                      gaacct  aagacctattatacagatattacgccagcgatagata


CAEEL-PGP-3____   679                N          AREEMIEEGAMEASLFDIFKYAS
                                                AR+ +IEEG MEASL+DI+ YA
                                     S          ARDNLIEEGGMEASLLDILCYAK
H_S09_0001706__12304 a            [agt]        gcgatagggggaggtccgacttga
                      gGTTAGCA  Intron 14  CAG cgaattaaggtaccttattgaca
                       <2-----[12306:13740]-2>tttccgcaacagatggcctaccca
                         13 n [13116:13178]


CAEEL-PGP-3____   703 P                        EMRNIIISLVFTLIRGFTWPAFSIV
                      P                        E+    I+L+F LIRG+TWP FSIV
                      P                        ELPMAGIALIFALIRGLTWPLFSIV
H_S09_0001706__13811 c                         gccaggagcatgtacgcatccttag
                      cGTAAGCG  Intron 15  CAGatctcgtctttcttggtcgcttctt
                      g<0-----[13814:14303]-0>atagttttgtccgatacgggatttc


CAEEL-PGP-3____   729 YGQLFK                         ILSAGGDDVSIKA-LLNSLW
                      YG+LF                          + S    +       + NS+
                      YGKLFL                         LFSNPDPNALANGNIFNSIC
H_S09_0001706__14379 tgattc                          tttacgcagcgagaatatat
                      agatttGTAAGGG  Intron 16  CAGttcacacactcagattactg
                      ccggcg<0-----[14397:15105]-0>gctcgtgtgacttcaccgct


CAEEL-PGP-3____   754 FILLAFTGGISTLISGSLLGKAGETMSGRLRMDVFR
                      F+LL +  GI+   SGSL G  GE ++ RLRMDVF+
                      FLLLGIGSGITAFASGSLFGITGEKVAMRLRMDVFK
H_S09_0001706__15166 tctcgagtgaagtgtgtttgaaggaggactcaggta
                      ttttgtgcgtcctccgcttgtcgaatctgtgtatta
                      cgatctatccaacataaactctcagaaggaagttcg


CAEEL-PGP-3____   790                          NIMQQDASYFDDSRHNVGSLTSRLAT
                                               NIM+QDASYFD+ +HN G+LT+ LA+
                                               NIMRQDASYFDNPKHNTGNLTAHLAS
H_S09_0001706__15274                           aaaccggtttgacacaagatagctgt
                      GTGAAAG  Intron 17  CAGattgaaccataacaaacgatccatcc
                      <0-----[15274:16212]-0>ccgcacccctttaactaatgattgaa


CAEEL-PGP-3____   816 DAPNVQA                                          822
                      D PNVQA
                      DTPNVQA
H_S09_0001706__16291 gacagcg                                        16311
                      accatac
                      cagtgag


H_S09_0001706__16312 GTAAGTA  Intron                                19077
                      <0-----[16312:19077]-0>
//

Gene 1
EXONS 381 16311
  Exon 381 514 phase 0
  Exon 811 912 phase 2
  Exon 1187 1256 phase 2
  Exon 2618 2742 phase 0
  Exon 6501 6556 phase 2
  Exon 8597 8667 phase 1

  ---

  Exon 8716 8835 phase 0
  Exon 8904 9035
       9036 9039 phase 0
  Exon 9103 9204 phase 0
  Exon 9779 9894 phase 0
  Exon 10588 10675 phase 2
  Exon 10735 10905 phase 0
  Exon 10993 11095 phase 0
  Exon 11962 12060 phase 1
  Exon 12155 12305 phase 1
  Exon 13741 13813 phase 2
  Exon 14304 14396 phase 0
  Exon 15106 15273 phase 0
  Exon 16213 16311 phase 0
//

Making a S in phase 2 intron
Making a S in phase 2 intron
Making a R in phase 2 intron
Making a V in phase 1 intron
Making a R in phase 2 intron
Making a G in phase 1 intron
Making a D in phase 1 intron
Making a S in phase 2 intron


>H_S09_0001706___HAECO_S09_Supercontig_0001706
 381-8667bp_AA
---
_8716-9035bp_AA
_9036-9039bp_Frameshift
_9040-16311bp_AA
_CAEEL-PGP-3_117-303AA---316-399AA_400AA_Frameshift_401-822AA

NASLYTLCERRIHSIRARYLRAVLRQDMTWLDQQQTGALTM
KMSSGMERIKDGIGDKLGLILASIGSFIGGISLGFYLSWRMTLVMLITVPLLIGATQFSG
KLLSRASKMENYAYSSAAALANEVIAGIRTVMAFNAQPFEIHRKIQWYDARRRSLNHGFL
EVLCCHLANCWESTECLNSLASGSL
```

```
FAVFWSVFLGTRRLSDVAP
QLGAFLGAKIAAADIFAVIDRVPEIDPMSNDGLTPEEFVGRLTFSNIHFTYPSRPTVEIL
DGVSF

VNPGETVALVGHSGCGKSTIISLLLRFYEQSAGMVALDGIPLRDYNVKWWRSVV
GVVQQEPVIFCATVAENVRMGDDSLTDEDVEEACKLANALGFINKLSEGFNTVIGEGAVQ
LSGGQKQRIAIARALVRNPQILLLDEATSALDTESERAVQEALDKARENRTTLCIAHRLS
TIRDSDKIIVFDEGHIVEQGTHDELMSIEDGVYRSMVKAQAIEKGEEDTTLDDVDPTEIR
RGVSRVVSEDDSQSRADRARESARLRQSMVSASTQEPEWEIESARDNLIEEGGMEASLLD
ILCYAKPELPMAGIALIFALIRGLTWPLFSIVYGKLFLLFSNPDPNALANGNIFNSICFL
LLGIGSGITAFASGSLFGITGEKVAMRLRMDVFKNIMRQDASYFDNPKHNTGNLTAHLAS
DTPNVQA
//

>H_S09_0001706___HAECO_S09_Supercontig_0001706
_381-8667bp
---
_8716-9035bp
_9036-9039bp_Frameshift
_9040-16311bp

aat
gcgagtctgtatacgctgtgcgagcgacggatacactccatcagagccagatatcttcga
gcagtactgcgtcaggacatgacatggttggatcaacaacaaaccggagcacttaccatg
aaaatgagcagtggaatggaacgaataaaggatggtatcggagacaagcttggactaatt
ctcgcatcgattggaagtttcatcggtggaatctcactcggtttctatttgagttggcgt
atgacgttggtaatgctcatcacggtaccattgctgataggagcaacgcaattttctgga
aagttactttcgcgtgcatctaaaatggaaaactacgcgtactcatcagcggcagccctt
gccaacgaagtgatcgccggaatacgaacggttatggccttcaacgctcaaccatttgag
attcaccggaagatacagtggtatgatgcacggcgccgtagcctgaatcatggatttcta
gaagttctctgctgtcatcttgccaattgttgggagtctacagaatgtttaaattcctta
gcatcaggcagctta

tttgctgtcttctggtcagtattcttgggtacacgacgtctgagtgatgtggctcca
caacttggagcattcctgggagcaaaaattgctgctgcagatatttttgctgttattgat
agggtgccagaaatcgatccgatgagcaacgacggtctaaccccagaggagtttgtagga
cgtctaaccttcagcaacattcatttcacttatccgtctagaccaacggtagagattctt
gacggtgtcagcttc

tgag

gttaaccctggtgaaactgtagctttggttgggcattcgggatgtggaaaatcgacgatc
atcagcttgctactccgattctacgagcaaagtgcaggaatggtggcccttgatggtatt
ccactgcgagattacaatgtgaaatggtggcgttcagttgttggtgttgtacaacaagaa
ccagtgatattctgtgccacggtcgctgaaaatgtgcgaatgggtgatgacagccttact
gatgaggacgttgaagaagcatgtaaacttgcaaacgctttgggcttcatcaataaattg
agtgagggattcaacactgtgatcggtgaaggagccgtgcagctatcaggcggccagaag
caacgaatcgctatagccagagccctggtcagaaaccctcaaattttgcttctggatgaa
gcgaccagtgctttggacacggaaagtgaacgtgctgtacaagaagcacttgacaaggct
cgtgaaaacaggactactctatgcatagctcacaggctttcgaccatcagggactctgac
aaaatcattgtattcgacgagggacatatagtcgaacaaggcacgcatgatgaactgatg
tcaattgaggatggtgtgtatcgtagcatggtgaaagcgcaagcgatcgaaaaaggcgaa
gaagacactaccctcgatgacgtggatccgacagaaatccggcgaggagtctcccgtgta
gtatcggaagacgacagtcaaagtcgtgcagatcgagccagagagtcagctagacttcgt
caaagcatggtcagcgcatcgacccaggaacctgaatgggaaattgaaagtgctcgtgac
aacttgatcgaagaaggcggaatggaagcttcgctgctcgacattctatgctacgccaaa
ccggaacttccaatggctggtattgctctgattttcgccttgatacgtggactcacgtgg
ccgctattttctattgtctacggcaagttgttcctgttgttctctaacccggatccgaat
gcgctagccaatggtaacatattcaactcgatctgtttcctgttacttggcattggatct
ggcatcacagcattcgcatctggatcattattcggtatcactggcgaaaaggtagcaatg
cggttacgaatggatgttttcaagaacatcatgcgccaagacgcctcctactttgataat
ccaaaacacaatacaggaaatttgacagctcatttggcatcagacacaccgaatgtgcaa
gcg
//
```

A.18 *HAECO*_S09_SUPERCONTIG_0059287 ≈ *CAEEL*-PGP-3

```
genewise
Query protein:       CAEEL-PGP-3___CAEEL-PGP-3 ZK455 7 1-1268AA
Target Sequence      H_S09_0059287___HAECO_S09_Supercontig_0059287

H_S09_0059287__    1                    CAG
                     <2-----[     :  433]-2>


CAEEL-PGP-3____  196  WQLTLVMMITVPLQLGSMYLSAK                      HLN
                      W++TLVM+I +PL +G+ +L+ K                       L+
                      WRMTLVMLILIPLVIGATHLTGK                      LLS
H_S09_0059287__  434  tcaatgacatactgaggactaga                       tct
                      ggtctttttttctttgccatcgaGTAAAAG  Intron 1   TAGttc
                     tgggggggccgaagaaaaatggag<0-----[504  : 1281]-0>gta


CAEEL-PGP-3____  222 RATKNEMSAYSNAGGMANEVIAGIRTVMAFNAQPFEINR        260
                     RA+K E  AYS+A  +ANEVIAGIRTVMAFNAQPFEI+R
                     RASKMENCAYSSAAALANEVIAGIRTVMAFNAQPFEIHR
H_S09_0059287__ 1291 cgtaagatgtttgggcgaggaggacagagtagcctgacc       1406
                     gccataagcacccctcaattcgtgcttctacactatag
                     gacgggtcgtaaagtcctagcctaagagccttgacatt


H_S09_0059287__ 1407 GTAAGCT  Intron                              2161
```

```
                                          <2-----[1407 :     ]-2>
//

Gene 1
EXONS 434 1406
  Exon 434 503 phase 0
  Exon 1282 1406 phase 0
//

>H_S09_0059287___HAECO_S09_Supercontig_0059287
_434-1406bp_AA
_CAEEL-PGP-3_196-260AA

WRMTLVMLILIPLVIGATHLTGKLLSRASKMENCAYSSAAALANEVIAGIRTVMAFNAQP
FEIHR
//

>H_S09_0059287___HAECO_S09_Supercontig_0059287

t
tggcggatgacgttggtgatgctcatcttgataccattggtaataggagcaacacatttg
acgggaaagttgctttcacgggcatccaagatggagaattgcgcgtattcatcagcagcg
gctctcgccaatgaagtgatcgccggtatacgaacggtaatggccttcaatgctcagcca
ttcgaaattcatcg
//
```

A.19 *HAECO*_S09_SUPERCONTIG_0023983 ≈ *CAEEL*-PGP-3

```
genewise
Query protein:       CAEEL-PGP-3____CAEEL-PGP-3 ZK455 7 1-1268AA
Target Sequence      H_S09_0023983___HAECO_S09_Supercontig_0023983

H_S09_0023983__    1          Intron     CAG
                     <0-----[     : 2572]-0>
                       10 n [1301 : 1310]


CAEEL-PGP-3____  823 AIDQRLAEVLTGIVSLFCGVGVAFYYGWNMAPIGLATALLL
                     AIDQRLAEVL G+ +L  G+ VAF +GWN+APIGLATAL+L
                     AIDQRLAEVLQGVCALVAGIAVAFSFGWNVAPIGLATALML
H_S09_0023983__ 2573 gagcccgggccggtgtgggagggtttgtaggcagcgagcat
                     ctaagtcattagtgcttcgtctctctggatcctgtcccttt
                     gttaacagggatttcgtttcatgtatcgtttcttgagtgga


CAEEL-PGP-3____  864                V          VVQSSVAQYLKFRGQRDMDSAIE
                                    V          V+QSSVAQYLKFRGQ+DMDSA+E
                                    V          VIQSSVAQYLKFRGQKDMDSAVE
H_S09_0023983__ 2696 g           [gtg]          gacttggcttatagcagagtggg
                      GTTAGTA  Intron 1   CAGtttacctcaatatggaaataccta
                      <1-----[2697 : 2926]-1>gatagtctgcgacaaagcgtatta


CAEEL-PGP-3____  888 ASR                                          890
                     AS+
                     ASQ
H_S09_0023983__ 2998 gac                                         3006
                     cga
                     ctg
```

```
H_S09_0023983__ 3007 GTAAGTA  Intron                             3085
                     <0-----[3007 :     ]-0>

//

Gene 1
EXONS 2573 3006
  Exon 2573 2696 phase 0
  Exon 2927 3006 phase 1
//

Making a V in phase 1 intron


>H_S09_0023983___HAECO_S09_Supercontig_0023983
_2573-3006bp_AA
_CAEEL-PGP-3_823-890AA

AIDQRLAEVLQGVCALVAGIAVAFSFGWNVAPIGLATALMLVVIQSSVAQYLKFRGQKDM
DSAVEASQ
//

>H_S09_0023983___HAECO_S09_Supercontig_0023983

gcgattgatcaacgactcgcagaggtgctgcaaggtgtttgtgccttggttgctggtatc
gcagttgcgttttcatttggctggaatgttgctcccattggtctggcaacggctctgatg
ttagtggtaattcaatcgtctgtcgctcagtacttgaaattcagaggacaaaaggacatg
gattcagctgttgaagccagtcag
//
```

A.20 *HAECO*_S09_SUPERCONTIG_0000863 ≈ *CAEEL*-PGP-3/4

```
genewise
Query protein:        CAEEL-PGP-4_1a CAEEL-PGP-4_F42E11_1a_1-1280AA
Comp Matrix:          blosum62.bla
Gap open:             12
Gap extension:        2
Start/End             local
Target Sequence       H_S09_0000863___HAECO_S09_Supercontig_0000863
Strand:               both
Start/End (protein)   local
Gene Paras:           worm.gf
Codon Table:          codon.table
Subs error:           1e-05
Indel error:          1e-05
Model splice?         model
Model codon bias?     flat
Model intron bias?    tied
Null model            syn
Algorithm             623

H_S09_0000863__   1          Intron      TAG                          244
                  <2-----[      :  244]-2>


CAEEL-PGP-4_1a_  943  LWQSLSFALAGSFFLWNFAIAYMFGLWLISNNWTTPFAVF
                      LWQ+LS+AL+ SF + NFAIAY FGLWLI N W+TPF VF
                      LWQALSLALSNSFVVVNFAIAYAFGLWLIRNGWSTPFIVF
H_S09_0000863__   245  ttcgctcgttaatgggatgagtgtgctcacagtaactagt
                      gtgactctctcagttttatctcactgtgttgagggcctttt
                      caggataaggattctgcctttcctttagtaatagtatcttc


CAEEL-PGP-4_1a_  983                          ---QVIEALNMASMSVMMAASYF
                                                VIEALNMASM+VMMAASYF
                                -:Q[caa]       LFLRVIEALNMASMTVMMAASYF
H_S09_0000863__  367 CAGTGAGTT  Intron 1   CAGAtttagaggtaagtaagaaggttt
                       <2-----[369  :  744]-2> tttgttactatcctctttcccat
                                              gtaagagggcggggtgggtcctc


CAEEL-PGP-4_1a_ 1003 PEYVRARISAGIMFTMIRQKAKIDNRGLTGETP
                      PEY+RARISAG+MFTM+RQ+ KIDN    GE P
                      PEYIRARISAGVMFTMMRQRPKIDNMSHQGEKP
H_S09_0000863__  815 cgtacgcatgggataaaccccaagaaaccggac
                      caatgcgtccgtttcttgagcataatgaagaac
                      cgtaaaacgttcgtaggaaagatccgctaaaaa


CAEEL-PGP-4_1a_ 1036                          DIRGDISMKGVYFAYPNRNRQLILNN
                                               ++GD++++ VYF+YP R RQL+L
                                              ALKGDVALRNVYFSYPARRRQLVLQG
H_S09_0000863__  914                          gcaggggcaagttttcgccccgccg
                      GTGAGTA  Intron 2   TAGctagatctgatatcaccgggatttag
                      <0-----[914  : 1011]-0>tcaacgagacttcgtgaatgatatac


CAEEL-PGP-4_1a_ 1062 FNMSAQFGETVALVGPSGCGKSTSIQLIERYYDAICGAVK      1101
                      +N+S + G+TVALVG SGCGKST IQL+ERYYDA+CG V
                      MNLSVRHGQTVALVGASGCGKSTVIQLVERYYDALCGTVV
H_S09_0000863__ 1090 aacagccgcaggtgggagtgaaagaccggcttggttgagga     1210
                      tatgtgagactcttgcggggagcttattagaaactggctt
                      gcaccatagtgaagacctcagcagagtcaaccttgtctga


H_S09_0000863__ 1211  GTTGCAG  Intron 3
                      <1-----[1211 :     ]-1>
                        10 n [1646 : 1655]
                       295 n [2811 : 3060]
                        10 n [3883 : 3892]
                        10 n [9704 : 9713]
                        10 n [19114:19123]


CAEEL-PGP-4_1a_ 1102 IDDHDIRDISVKHLRHNIALVGQEPTL                  1128
H_S09_0000863__      ---------------------------


CAEEL-PGP-4_1a_ 1129 FNLTI              R          ENITYGLENVSQEQ
                      FN+TI              R          ENI YGL+  SQE+
                      FNVTI              R          ENIMYGLDKCSQEE
H_S09_0000863__19124 tagaac          [cgc]        gaaatgcgattcgg
                      tatctgGTACGTT  Intron 4   CAG aattagtaagcaaa
                      cctaa <2-----[19141:19623]-2>cgccgccgtgtcaga


CAEEL-PGP-4_1a_ 1149 VEKAATLANIHSFVENLPE                       GYDTSVG
                      +  AA LANIH F+ +LPE                       Y+T VG
                      IVHAARLANIHDFIASLPE                      EYNTVVG
H_S09_0000863__19667 agcggccgaacgtagaccg                       gtaaggg
                      ttaccgtcataattcgtcaGTTTGGG  Intron 5   TAGaaacttg
                      aataatcattcccattttg<0-----[19724:19879]-0>gctgccg
```

```
CAEEL-PGP-4_1a_ 1175 ASGGRLSGGQKQRIAIARAIVRNPKILLLDEATSALDTESEK
                      A GG LSGGQKQRIAIARAIVR+PKILLLDEATSALDTESEK
                      AKGGLLSGGQKQRIAIARAIVRDPKILLLDEATSALDTESEK
H_S09_0000863__19901 gaggtctggcaccagagcgagagcaacccgggaagcgagaga
                      caggttcggaaagtctcgcttgacattttaaccgctacagaa
                      agtgagtataggaacttttccatagagttcatttggcaatag


CAEEL-PGP-4_1a_ 1217                          IVQEALDKARLGRTCVVIAHR
                                              +VQEALD+ARLGRTC+VIAHR
                                              VVQEALDRARLGRTCLVIAHR
H_S09_0000863__20027                          ggcggtgcgccgcatcgagcc
                      GTAAGTT  Intron 6   CAGttaactagcgtggcgtttcag
                      <0-----[20027:20349]-0>ggaatataaagaattatatta


CAEEL-PGP-4_1a_ 1238 LSTIQ                                        1242
                      LSTIQ
                      LSTIQ
H_S09_0000863__20413 ttaac                                      20427
                      tccta
                      attag


H_S09_0000863__20428 GTTTGTC  Intron                            21232
                      <0-----[20428:     ]-0>
//

Gene 1
EXONS 245 20427
  Exon 245 368 phase 0
  Exon 745 913 phase 2
  Exon 1012 1210 phase 0

  ---

  Exon 19124 19140 phase 1
  Exon 19624 19723 phase 2
  Exon 19880 20026 phase 0
  Exon 20350 20427 phase 0
//

Making a Q in phase 2 intron
Making a R in phase 2 intron


>H_S09_0000863___HAECO_S09_Supercontig_0000863
_245-1210bp_AA
---
_19124-20427bp_AA
_CAEEL-PGP-3_OR_4_CAEEL-PGP-4_1a_943-1101AA---1129-1242AA

LWQALSLALSNSFVVVNFAIAYAFGLWLIRNGWSTPFIVFQLFLRVIEALNMASMTVMM
AASYFPEYIRARISAGVMFTMMRQRPKIDNMSHQGEKPALKGDVALRNVYFSYPARRRQL
VLQGMNLSVRHGQTVALVGASGCGKSTVIQLVERYYDALCGTVV


FNVTIRENIMYGLDKCSQEEIVHAARLANIHDFIASLPEEYNTVVGAKG
GLLSGGQKQRIAIARAIVRDPKILLLDEATSALDTESEKVVQEALDRARLGRTCLVIAHR
LSTIQ
//

>H_S09_0000863___HAECO_S09_Supercontig_0000863
_245-1210bp
---
_19124-20427bp

gcttatggcaggcactttcactagcgttgtcaaatagtttcgttgtggtcaactttgct
attgcctacgcttttggtctatggcttatacgaaatggatggagtacacctttcattgtt
ttccaattgtttttaagagtgatagaggcgttgaacatggcgtcgatgactgtgatgatg
gctgcctcctatttccccgagtatatacgagcacgaatctcggctggtgtcatgtttaca
atgatgcgacaacgaccgaaaattgacaacatgagccatcaaggagaaaaaccagctctc
aaaggagacgtggcactgagaaacgtttatttctcgtatccggcacgacgtcggcaacctt
gtacttcaaggcatgaacctaagcgtccgacatggacagactgtggcattagtgggagcc
agcggttgcggaaagagcacagtgatacagcttgtcgaacgatactacgatgctttgtgt
ggcactgtggtaa


ttcaacgttacaatacgcgagaacatc
atgtacggcctggataagtgttcccaagaggaaatagtacatgcagcacgtctcgcaaat
attcacgacttcatagctagtcttcctgaggagtacaatacggtcgtcggggcaaagggt
gggttactgtctggaggtcaaaagcagcgaatagccattgctcgtgctatcgtcagagat
ccaaagatactgcttcttgacgaagctactagtgcgctggacacagaaagtgaaaaggtg
gtgcaagaagctttagatcgagcacgactgggacgaacttgtctagttatagctcatcga
ttatctactatacag
```

A.21 *HAECO*_S09_SUPERCONTIG_0001880 ≈ *CAEEL*-PGP-3/4

```
genewise
Query protein:       CAEEL-PGP-4_1a CAEEL-PGP-4_F42E11_1a 1-1280AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            global
Target Sequence      H_S09_0001880___HAECO_S09_Supercontig_0001880
                                    _1-30001bp

Strand:              both
Start/End (protein)  global
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623
```

```
H_S09_0001880__      1          Intron     CAG
                     <0-----[     :13980]-0>
                      449 n [5271 : 5719]
                      234 n [7717 : 7950]
                       10 n [9514 : 9523]
                      963 n [10866:11828]
                       10 n [12734:12743]


CAEEL-PGP-4_1a_     47 FRHSGCADYLLLLGGLVLSAANGALLPFNSLIFEGITNVLMK
                       FR++   D+ L+L G V +A  G     +SL+F  + + L+
                       FRYATKLDFCLMLLGAVFAATQGTFNSVSSLVFRHLMDALII
H_S09_0001880__13981 tctgaacgtttatcgggtggacgatatgttcgtcccaggcaa
                      tgaccatatgttttgcttcccagctactcctttgattacttt
                      tctaaattctggagaggtactaatcttaatgttgttgttaca


CAEEL-PGP-4_1a_     89                G          EAQWQNGTFDYDTFSSGIQHYCL
                                      G          E +WQ G FD   F+    +
                                      G          EFEWQAGIFDDYEFTQLAMNSVY
H_S09_0001880__14107 g           [ggt]          gtgtcggatggtgtaccgaatgt
                      GTGGAAC  Intron 1   TAGgatagacgttaaaatcatctacta
                      <1-----[14108:14173]-1>tacagatcaccttgcgagtgttct


CAEEL-PGP-4_1a_    113 LYFLLGVLMFTCTYFS                          NACLFTMAER
                       Y L G+++FT  ++S                           C  T+ ER
                       RYTLFGLIQFTLGFLS                          MCCWHTVCER
H_S09_0001880__14245 atactgcactacgtct                           atttcagtgc
                      gacttgttatctgttcGTCAGTT  Intron 2   TAGtgggactgag
                      actattcaacttatag<0-----[14293:14794]-0>gttgtaataa


CAEEL-PGP-4_1a_    139 RLYCIRKHLLQSVLRQDAKWFDENTVGGLTQKMS
                       ++Y IR     SV+RQD  WFD+N  G LT +MS
                       QVYQIRNRFFGSVIRQDMAWFDQNDSGALTTRMS
H_S09_0001880__14825 cgtcaaacttgtgaccgagttgcagaggcaacaa
                      ataatgagttgcttgaatcgtaaaaggctccgtg
                      aatacatcttaagaagtgcgtcattcatgatggc


CAEEL-PGP-4_1a_    173                S          GIEKIKDGIGDKIGVLVSGIATF
                                      GI++I+DGIGDK+  + + +ATF
                                      D          GIDRIRDGIGDKLDAMFAYFATF
H_S09_0001880__14927 g           [gat]          gagcacggaggacggatgttgat
                      aGTATGTT  Intron 3   CAG gtagtgagtgaatacttcatcct
                       <2-----[14929:15689]-2>taatgagttcccacttgcgttcgc


CAEEL-PGP-4_1a_    197 ISGVALGFYM               C          WQLTLVMLVTVPL
                       I+G+ + +                            WQ+TLVM+   P+
                       IAGITVALSC               S          WQMTLVMIGFFPI
H_S09_0001880__15760 aggaaggcata          [agt]          tcaacgaagttca
                      tcgtctctgggGTAAGAA  Intron 4   CAG gatcttttgttct
                      ctaaattgtt <2-----[15792:15905]-2>tgagtgtgtgtcac


CAEEL-PGP-4_1a_    221 QLGSMYLSAK                          HLNRATKNEMSAYSSA
                         G + +++                           + +    E+  Y  A
                       FFGPLTVTSM                          IMGKVVPKEQEFYVRA
H_S09_0001880__15946 ttgccagata                          aagaggcagcgttgcg
                      ttgctctcctGTGAGGG  Intron 5   CAGttgattcaaaatatgc
                      ctcactatag<0-----[15976:16029]-0>tgagcacagagtctgg


CAEEL-PGP-4_1a_    247 GGMANEVIAGIRTVIAFNAQPFEI                      270
                       G  A EV+ GIRTV+AFN Q  EI
                       GSTAAEVVNGIRTVVAFNGQEKEI
H_S09_0001880__16078 gtaggggagacagggtagcgaga                    16149
                      gcccaattagtgcttctagaaagt
                      ctctagacccactgacttaggagt
                                    884 n [16150:17033]


CAEEL-PGP-4_1a_    271 ERYGAQLAKARKMGIRKAIVLALCSAMPLFLMFVLMAGAFWYGAILTSYG
H_S09_0001880__      ------------------------------------------------------


CAEEL-PGP-4_1a_    321 VATSGTTFGVFWAVILGTRRLGEAAPHMGAITGARLAVNDIFKVIDHEPE
H_S09_0001880__      ------------------------------------------------------


CAEEL-PGP-4_1a_    371 INCTKQEGRRPDKVNGKLVFDNIQFTYPTRPDVKILKGVSFEVNPGETIA
H_S09_0001880__      ------------------------------------------------------


CAEEL-PGP-4_1a_    421 LVGHSGCGKSTSIGLLLMRFYNQCAGSIKLDGIPIEDYNIQWLRSTIGIVQ
H_S09_0001880__      ------------------------------------------------------


CAEEL-PGP-4_1a_    471 QEPIIFLATVAENVRMGDDSITDKDIENACRQANAHDFIGKLSE    514
H_S09_0001880__      ----------------------------------------------


H_S09_0001880__          Intron     CAG                          17518
                         <0-----[     :17518]-0>


CAEEL-PGP-4_1a_    515 GYNTVIGAGAVQLSGGQKQRVAIARAIVRKPQILLLDEATSALDTE
                       GY T IG G VQLSGGQKQRVAIARA+VR P+ILLLDEATSALD E
                       GYQTKIGEGGVQLSGGQKQRVAIARALVRNPRILLLDEATSALDAE
H_S09_0001880__17519 gtcaaaggggcttggcaccggagcgcgaaccacccgggaagtggg
                      gaacatgaggtatcggaaagtctcgcttgacgttttaaccgctaca
                      ccgaaaaatatgactaaggattctacgtgtaaatagtgagttgtaa


CAEEL-PGP-4_1a_    561 SERMVQTALDK                          ASEGRTTLCIAHRLS
                       SE +VQ AL+                           A  GRTT+ IAHRLS
                       SESIVQQALEN                          AQSGRTTISIAHRLS
H_S09_0001880__17657 agaagccgcga                           gctgaaaatagcact
                      gagttaactaaGTACGTT  Intron 7   CAGcacggcctctcagtc
                      cattcaggtat<0-----[17690:18028]-0>tacggcgtgcatatg


CAEEL-PGP-4_1a_    587 TIRNASKILVFDQGLIPER                G          IHDQ
                       TI+N  +I VF+ G I E                 G           HD+
                       TIKNVDRIYVFNNGRIVED                G          NHDE
H_S09_0001880__18074 aaaaggcatgtaagaagggg             [gga]        acgg
                      ctaatagtattaaggttaa GTGAGCT  Intron 8   TAGgaaaa
                      aaatctcccaccttgagac <1-----[18132:18212]-1>accca


CAEEL-PGP-4_1a_    611 LIRQNGIYANMVRAQEIEKAKDDTTQDDDELV
                       L++ NG+Y+ +VRAQEIE+     +  DDE
```

```
                     LMKMNGLYSELVRAQEIEQL--EKSGSDDE-T
H_S09_0001880__18227 caaaagtttgtgagcgagcc  gatgaggg a
                     ttatagtacattgcaataat  aacggaaa c
                     cgagcggcggggagagtgat  gatgccta t


CAEEL-PGP-4_1a_  643               E          EDNYSISRRLSTSEEELRKSKSL
                                   E N ++ RR          SK L
                                   A          EHNVTLMRR---------RSKRL
H_S09_0001880__18314 g          [gct]         gcagacaaa          aaacc
                      GTGAGAG  Intron 9   CAGcaaatcttgg          ggagt
                      <1-----[18315:18383]-1>tactgaggga          acgcc


CAEEL-PGP-4_1a_  667 LRDSTRFSQSMLSVTSQVPDWEMESAR
                      R  +R         +++   E+E+
                     SRSISR--------PTELRGQELENLE
H_S09_0001880__18428 tctatc         cagccgcgtgacg
                     cgctcg         ccatggaataataGTAACCT  Intron 10
                     ctccat         cgatttagaaccg<0-----[18485:18570]


CAEEL-PGP-4_1a_  694    EEMIEEGAMEASMMDIFRFAKPEKMNIVIALIFTLIRGITWPAFSV
                        EE+ E+    AS++DI++FA+ E   + +ALI+ + RG+T+P FS+
                        EEVEEKKVKGASLLDILKFARQEWCQLAVALILAVARGMTFPVFSI
H_S09_0001880__18568    gggggaagaggaccgacatgacgttccgggcacgggcgaatcgtta
                     TAGaataaaatagcgttattatcgaaggatctctttctcggtctcttct
                     -0>aacggaagaaattattggtaagagcagagctcttattgggctgtga


CAEEL-PGP-4_1a_  740 VYGQLFKVFAEGGEDLPVNA              -          ---
                     +YGQ+FKV A      P+N
                     IYGQMFKVRAFRRLLFPMNY            G          NSQ
H_S09_0001880__18709 atgcatagagtactctcaatg           [ggc]        atc
                     tagattatgctggtttctaagGTTAATT  Intron 11  CAG aca
                     tttagcggaacgtggctgct <2-----[18771:19450]-2>ccca


CAEEL-PGP-4_1a_  760 -----LISSLWFVLLAVTSAVTTFISGSLLGKTGETMSSRLRMDVFKNI
                           ++++WF LL ++S ++T+ISG L G+ GE++++RLR+ +F NI
                     KLHGATMNAIWFSLLGISSGISTMISGYLFGRIGESLTNRLRLSLFTNI
H_S09_0001880__19461 accggaaagatttccgatagaaaaatgtctgaaggtcaaaccctctaaa
                     atagcctactgtcttgtcggtgcttcgattggtgactcagtgtcttcat
                     gatacggtgagcaagcttttacagttctgtgacaatataataacgtgtc


CAEEL-PGP-4_1a_  804 MQQDAT              -          YFDDPKHNVGNLTSRLA
                     ++Q +                         YFD   H  G LT+RLA
                     VKQVSK             E          YFDHEDHASGKLTTRLA
H_S09_0001880__19608 gacgaag          [gag]        ttgcggcgtgataaatg
                     taatga GTCAGAA  Intron 12  GAGaataaaaaccgatccgtc
                     agaacg <1-----[19627:19682]-1>gctccatttaaagcaagg


CAEEL-PGP-4_1a_  827 TDSQNVQA                        AIDHRLAEVLNGVVSLFT
                     TD+ N++A                        AID RLA+V++ V S++
                     TDAPNIRA                        AIDQRLADVVSAVSSIIG
H_S09_0001880__19736 aggcaaag                        gagcccgggtggttaag
                     caccatgcGTATGGT  Intron 13  TAGctaagtcattcctccttg
                     gcaattag<0-----[19760:20114]-0>actattttccatgggata
                              10 n [20086:20095]


CAEEL-PGP-4_1a_  853 GIAVAFWFGWSMAPIGLITAL                          LLVIA
                     GI++AF +G +MAPIG++TA+                           L+
                     GISIAFSYGPAMAPIGVLTAV                          TLITL
H_S09_0001880__20169 gatagtttgcgagcaggcagg                          acaac
                     gtctctcagcctcctgttcctGTAAGAA  Intron 14  CAGcttct
                     ctcaatccataggatactgca<0-----[20232:20289]-0>tacag


CAEEL-PGP-4_1a_  879 QSAVAQYLKYRGPKDMESAIEASR                        IV
                     Q+ VA+YLK RG +D   A E SR                        +
                     QTLVARYLKVRGQRDAVLAEEPSR                        LA
H_S09_0001880__20305 cacggctcagcgccgggcgggcta                        tg
                     acttcgatatggagacttcaaccgGTAAGCC  Intron 15  CAGtc
                     aattcatggcaaattagaagagaa<0-----[20377:20442]-0>gt


CAEEL-PGP-4_1a_  905 TESISNWKTVQALTKQEYMFHAFTAASKNPRKRAFTK
                     TE+I   KTVQ LTK+  ++  F     P KRA+ +
                     TEAIEQHKTVQYLTKERQFLDKFVTQMHGPHKRAIFR
H_S09_0001880__20449 aggagccaagctcaagccttgatgacacgccaagatc
                     cactaaaactaatcaagattaattcatagcaagcttg
                     gagaagtgacacatgaggcgtactcagtttcagaata


CAEEL-PGP-4_1a_  942               G          LWQSLSFALAGSFFLWNFAIAYM
                                  G          + QSL++AL+ SF   NFAIAY+
                                  G          IVQSLTYALSVSFVNLNFAIAYL
H_S09_0001880__20560 g          [ggt]        agcttatgctgatgatatgagtc
                      GTACTGA  Intron 16  CAGgttactcactctgttatatctcat
                      <1-----[20561:20880]-1>tttgagttccctttccgctcccct


CAEEL-PGP-4_1a_  966 FGLWLISNNWTTPFAVF              Q          VIEALN
                     +G+WL+     +P+ VF              Q          VIE+LN
                     YGIWLVGRRICSPYTVF              Q          VIESLN
H_S09_0001880__20952 tgatcggaaatttctagtc          [cag]        gagtta
                     agtgttgggtgccacttaGTAAGTT  Intron 17  CAG ttacta
                     ctcggtggacctgtggc <2-----[21005:21063]-2>ggtacgc


CAEEL-PGP-4_1a_  990 MASMSVMMAASYFPEYVRARISAGIMFTMIRQKAKIDNRGLTGETP
                      ASMS++  A+YFPEYVRAR+SA ++F M+R K KID+    G
                     TASMSLIAFATYFPEYVRARLSAALLFRMLRDKPKIDSLSPLGMQT
H_S09_0001880__21083 agtatcagtgattcgtgcgcctggcctcacagacaagactctgaca
                     ccctcttctccatcaatgcgtccctttgttgaacatagtcctgtac
                     cgggtgtatctcctgtcttaaaacttcagtgcagacccgcaaagat


CAEEL-PGP-4_1a_ 1036                    DIRGDISMKGVYFAYPNRN
                                         +RG I +  + F+YP
                                        KLRGSIHFSDLSFSYPVSR
H_S09_0001880__21221                    atcgtacttgcttttcgac
                     GTGAGTT  Intron 18  CAGatggctatcatctcactgg
                     <0-----[21221:21276]-0>gacaatctttgccatcctt


CAEEL-PGP-4_1a_ 1055 RQLILNNFNMSAQFGET                          1071
                     R ++L  + +    GE+
                     RDMVLKGITLKVITGES
H_S09_0001880__21334 agagcagaacagaaggta                        21385
                     gatttagtctattcgac
                     atgacaacttgattcac


H_S09_0001880__21386 GTACGAA  Intron
                     <1-----[21386:     ]-1>
                      480 n [21872:22351]


CAEEL-PGP-4_1a_ 1072 VALVGPSGCGKSTSIQLIERYYDAICGAVKIDDHDIRDISVKHLRHNIAL
H_S09_0001880__      ------------------------------------------------------


CAEEL-PGP-4_1a_ 1122 VGQEPTLFNLTIRENITYGLENVSQEQVEKAATLANIHSFVENLPE1167
H_S09_0001880__      ----------------------------------------------


CAEEL-PGP-4_1a_ 1168  GYDTSVGASGGRLSGG
                      GYDT VG  G  LSGG
                      GYDTIVGERGSMLSGG
H_S09_0001880__22353  gtgaagggcgaattgg
```

```
                    gaacttgagggttcgg
                    ctcgtttgactggaca

CAEEL-PGP-4_1a_ 1184 QKQRIAIARAIVRNPKILLLDEATSALDTESEK
                     QKQRIAIARA++R+PKILLLDEATSALDTESEK
                     QKQRIAIARAVIRDPKILLLDEATSALDTESEK
H_S09_0001880__22401 caccagagcggacgcaatctgggaagtgagaga
                     aaagtctcgcttgacatttaaccgctacagaa
                     gagaccaaattcttaatacatatgtgattataa


CAEEL-PGP-4_1a_ 1217                          IVQEALDKARLGRTCVVIAHRLSTIQ
                                              IVQEAL+KAR GRTC+VIAHRLS+IQ
                                              IVQEALEKARQGRTCIVIAHRLSSIQ
H_S09_0001880__22500                          agcggcgagacgcatagagcccttac
                     GTACGTC  Intron 20  CAGttaactaacgaggcgtttcagtccta
                     <0-----[22500:22554]-0>taaaggaagaatatttccttagttta


CAEEL-PGP-4_1a_ 1243 NADKIIVCRNGKAI                          EEGTHQTLLARR
                     NAD IIV ++G                              E+GTHQ LLAR
                     NADLIIVIKDGMVE                           EQGTHQQLLARE
H_S09_0001880__22633 aggcaagaaggagg                           gcgaccctcgag
                     acatttttaagttaGTAAGCT  Intron 21  CAGaagcaaattcga
                     tattccctacagtg<0-----[22675:23148]-0>gagacgaactaa


CAEEL-PGP-4_1a_ 1269 GLYYRLVEKQST-                                    1280
                     GLY  +V KQ
                     GLYASMVTKQDLK
H_S09_0001880__23185 gctgaagaacgca                                   23223
                     gtacgttcaaata
                     catacgacaatcg


CAEEL-PGP-4_1a_      *
                     *
                     *
H_S09_0001880__23224 t                                               23226
                     a
                     a


H_S09_0001880__23227       [23227:29424]                             29424
                     10 n [23649:23658]
                     10 n [28765:28774]
                      1 n [28839:28839]
//

Gene 1
EXONS 13981 23223
  Exon 13981 14107 phase 0
  Exon 14174 14292 phase 1
  Exon 14795 14928 phase 0
  Exon 15690 15791 phase 2
  Exon 15906 15975 phase 2
  Exon 16030 16149 phase 0

  ---

  Exon 17519 17689 phase 0
  Exon 18029 18131 phase 0
  Exon 18213 18314 phase 1
  Exon 18384 18484 phase 1
  Exon 18571 18770 phase 0
  Exon 19451 19626 phase 2
  Exon 19683 19759 phase 1
  Exon 20115 20231 phase 0
  Exon 20290 20376 phase 0
  Exon 20443 20560 phase 0
  Exon 20881 21004 phase 1
  Exon 21064 21220 phase 2
  Exon 21277 21385 phase 0

  ---

  Exon 22353 22499 phase 2
  Exon 22555 22674 phase 0
  Exon 23149 23223 phase 0
//

Making a G in phase 1 intron
Making a D in phase 2 intron
Making a S in phase 2 intron
Making a G in phase 1 intron
Making a A in phase 1 intron
Making a G in phase 2 intron
Making a E in phase 1 intron
Making a G in phase 1 intron
Making a Q in phase 2 intron


>H_S09_0001880___HAECO_S09_Supercontig_0001880
 13981-16149bp_AA
---
_17519-21385bp_AA
---
_22353-23223bp_AA
_CAEEL-PGP-3_OR_4_CAEEL-PGP-4_1a_47-270AA---515-1071AA---1168-1280AA

FRYATKLDFCLMLLGAVFA
ATQGTFNSVSSLVFRHLMDALIIGEFEWQAGIFDDYEFTQLAMNSVYRYTLFGLIQFTLG
FLSMCCWHTVCERQVYQIRNRFFGSVIRQDMAWFDQNDSGALTTRMSDGIDRIRDGIGDK
LDAMFAYFATFIAGITVALSCSWQMTLVMIGFFPIFFGPLTVTSMIMGKVVPKEQEFYVR
AGSTAEEVVNGIRTVVAFNGQEKEI


GYQTKI
GEGGVQLSGGQKQRVAIARALVRNPRILLLDEATSALDAESESIVQQALENAQSGRTTIS
IAHRLSTIKNVDRIYVFNNGRIVEDGNHDELMKMNGLYSELVRAQEIEQLEKSGSDDETA
EHNVTLMRRRSKRLSRSISRPTELRGQELENLEEEVEEKKVKGASLLDILKFARQEWCQL
AVALILAVARGMTFPVFSIIYGQMFKVRAFRRLLFPMNYGNSQKLHGATMNAIWFSLLGI
SSGISTMISGYLFGRIGESLTNRLRLSLFTNIVKQVSKEYFDHEDHASGKLTTRLATDAP
NIRAAIDQRLADVVSAVSSIIGGISIAFSYGPAMAPIGVLTAVTLITLQTLVARYLKVRG
QRDAVLAEEPSRLATEAIEQHKTVQYLTKERQFLDKFVTQMHGPHKRAIFRGIVQSLTYA
LSVSFVNLNFAIAYLYGIWLVGRRICSPYTVFQVIESLNTASMSLIAFATYFPEYVRARL
SAALLFRMLRDKPKIDSLSPLGMQTKLRGSIHFSDLSFSYPVSRRDMVLKG
ITLKVITGES


GYDTIVGERGSMLSGGQKQRIAIARAVIRDPKILLLDEATS
ALDTESEKIVQEALEKARQGRTCIVIAHRLSSIQNADLIIVIKDGMVEEQGTHQQLLARE
GLYASMVTKQDLK
//

>H_S09_0001880___HAECO_S09_Supercontig_0001880
_13981-16149bp
---
_17519-21385bp
---
_22353-23223bp

tttcgctatgcaacaaaacttgatttctgtttgatgttactgggagcggtgtttgcagcc
actcaaggaactttcaatttctgtatcatctctggtttttcggcatcttatggatgctcta
atcataggtgaattcgaatggcaagctggcatattcgacgattatgagttcacgcaactg
gctatgaattctgtctatagatacactctatttggtctcatacaattcactcttggattt
ctatcgatgtgttgttggcataccagtatgtgaacgacaagtatatcaaatcagaaatcgc
ttttttggatcagtgatacgacaggatatggcctggtttgaccaaaatgatagcggagct
ctgacaactcggatgagcgatggaatagatcggatacgggatggtatcggcgacaaactc
gatgctatgttcgcgtattttgccacgttcatcgctggaataacagttgctctgagttgt
agttggcaaatgactctggttatgattgggtttttcccaatcttctttggcccactcact
gtaacttcaatgattatgggaaaggtcgtacccaaagagcaagagttttacgttcgggcg
ggctctaccgctgaagaggtagtcaacggcatacgcactgtggtagcctttaatggacag
gagaaagagatt


ggctaccagacaaaaatagggagaggtggagttcagttatccggtggacaaaagcag
cgagttgctatcgctcgagccctggttaggaatccacgaatacttctactggatgaggca
acgagtgctttggatgcagaaagcgaaagtattgtccaacaggcgcttgaaaatgctcaa
tccgggaggaccacgatttcgatcgcacatagactttcgacaataaaaaatgtcgatcgc
atctacgtattcaacaatggtaggatagtggaagacggaaacccacgacgaactcatgaaa
atgaacgggttgtactcggagttggtgagagcgcaagagattgagcaacttgagaaatct
gggagcgacgatgaaactgctgaacacaatgtgacactgatgaggagaagaagcaagcgc
ctctcccgttccatctcacgtcccacggaacttcgtggtcaagagttagaaaacctcgag
gaagaagtcgaggagaaaaaagtgaaaggagcaagtcttctagatattctgaagtttgca
agacaggaatggtgccaactggcagtggcccttatccttgctgtagctcgtgggatgacg
ttccctgttgttttcgataatttatggtcaaatgttcaaggtgagagcattcaggcgtttg
ctgttccctatgaactatggcaactcccaaaagctacatggagccacgatgaatgcgata
tggttctcactactgggcatttctagtggtataagcacaatgatttctggctatctgttt
gggagaatcggagaatctctaacaaatagactacgtctatccctgtttacgaatatcgta
aagcaagtaagcaaggagtactttgaccacgaagatcatgcttcaggaaaattgaccaca
agattggcgacggacgcaccaaatattagagcggcaatcgatcaacgtcttgctgatgtc
gtctcagctgtgtcgtcgataattggagggcatttccatagcatttttcctacggacctgca
atggcgccaatggagtcctacggccgtaactctaatcacactgcaaacacttgttgccc
cgatatctgaaggtccgaggacaacgtgatgcagtgctagcagaggaaccgtcaagattg
gctacggaagcgatagaacagcataagacagtccaatacctaactaaggaacggcagttc
ttggataaattcgttacccaaatgcatggtcctcacaaaagggcaatatttcgaggtatt
gttcagtcattgacttatgccctctccgttagttttgtcaacttgaacttgccatcgcc
taccttacggtatctggctggttgggaggagaatctgctctccgtatacggtgttccag
gtgattgaatccttgaacaccgcgtcgatgtctctgattgcatttgccacttacttccct
gagtatgtccgtgctcgactatcagcagccctttttttccgaatgcttagggacaaaccg
aaaatcgacagcctgtccccattaggaatgcaaactaagttacgcggatcaattcacttt
tctgatctgtccttctcatatcccgtcagtcgtagagatatggtactcaaagga
atcactcttaaggtaattactggcgaatcca


ggctatgacacgattgttggtgagcgaggcagtatgttgtca
ggcggacagaaacagcgaatcgccatagcacggagctgttatccgtgatccaaaaatttta
ctcttagatgaagctacgagtgcgttagatactgaaagtgaaaaaattgtacaagaagcg
ctggaaaaagcgagacaaggtcgaacttgtattgtcatcgctcatcgactgtcttctatt
caaaatgcagatcttatcatcgtcattaaagacggaatggttgaggagcaagggacacac
cagcaattactcgctagagaaggcctatatgcaagcatggtaaccaaacaagatctcaag
//
genewise
Query protein:       CAEEL-PGP-4_1a_CAEEL-PGP-4_F42E11_1a_1-1280AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0001880___HAECO_S09_Supercontig_0001880
                                    _23227-30001bp
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623
H_S09_0001880__ 23227          Intron     CAG
                     <2-----[     :29424]-2>
                       10 n [23649:23658]
                       10 n [28765:28774]
                        1 n [28839:28839]


CAEEL-PGP-4_1a_ 1070  ETVALVGPSGCGKSTSIQLIERYYDAICGA                1099
                      +TVALVGPSGCGKSTSIQLIER+YD + G+
                      KTVALVGPSGCGKSTSIQLIERFYDPVAGS
H_S09_0001880__29425  aaggtggcagtgaaatacacagcttgcgggt               29516
                      gactcttgcggggagcctattagtaactcgc
                      gggtagtatcttaactttatcaaccttgtat


CAEEL-PGP-4_1a_ 1100 V                                              1100
                     !!
H_S09_0001880__29517 gt                                            29518


CAEEL-PGP-4_1a_ 1101                K          IDDHDIRDISVKHLRHNIALVGQ
                                               +D+ D R+++++HLR  ++LVGQ
                                    A          FDEVDARELNLRHLRSQMSLVGQ
H_S09_0001880__29519 g          [gct]         tgggggcgcataccctcatcggc
                      GTGAGGC  Intron 1   CAGctaatacgatatgatgcatcttga
                      <1-----[29520:29750]-1>tttaacttagtggtctaagatcaa


CAEEL-PGP-4_1a_ 1125 EPTLFNLTI              -          ---RENITYGLENV
                     EP L  +                           RENI YGLE
                     EPILSTIPS              E          NFCRENIAYGLEQA
H_S09_0001880__29822 gcactaactg           [gaa]        attagaagtgcgcg
                     acttcctcc GTACGAA  Intron 2   CAGaatggaatcagtaac
                     atttattaa <1-----[29850:29899]-1>acttaactgcccgag


CAEEL-PGP-4_1a_ 1145 SQEQVEKAATLANIHSFVE                           1163
                     + +Q+E AA LAN H+F+
                     TVDQIENAAKLANAHNFII
H_S09_0001880__29944 aggcagaggatgagcataaa                          30001
                     ctaataaccatcacaattt
                     tttgtatatgggtcccccc
//

Gene 1
EXONS 29425 30001
  Exon 29425 29516
       29517 29518
       29519 29519 phase 0
  Exon 29751 29849 phase 0
  Exon 29900 30001 phase 1
//

Making a A in phase 1 intron
Making a E in phase 1 intron


>H_S09_0001880___HAECO_S09_Supercontig_0001880
_29425-29516bp_AA
_29517-29518bp_Frameshift
_29519-30001bp_AA
_CAEEL-PGP-3_OR_4_CAEEL-PGP-4_1a_1070-1099AA_1100AA_Frameshift_1101-
```

1163AA

KTVALVGPSGCGKSTSIQLIERFYDPVAGS

AFDEVDARELNLRHLRSQMSLVGQEPIL
STIPSENFCRENIAYGLEQATVDQIENAAKLANAHNFII
//

>H_S09_0001880___HAECO_S09_Supercontig_0001880
_29425-29516bp
_29517-29518bp_Frameshift
_29519-30001bp

ggaagacggttgcattggttggacctagcggttgtggaaaaagcacttctattcaactt
atcgaacgattctacgatcctgtggctggatct

gt

gct
tttgatgaagtagacgctcgtgaactgaatttgaggcatctccgttcacaaatgtcactt
gtcggacaagaacctattctttcaactattccatcagaaaacttttgtagagaaaacatt
gcgtacggcctcgagcaagcgactgttgatcagattgaaaatgcagctaagttggcgaat
gcccacaacttcatcatca
//

A.22 *HAECO*_S09_SUPERCONTIG_0057179 ≈ *CAEEL*-PGP-4

```
genewise
Query protein:      CAEEL-PGP-4_1a CAEEL-PGP-4 F42E11 1a 1-1280AA
Target Sequence     H_S09_0057179___HAECO_S09_Supercontig_0057179

H_S09_0057179__-1743          Intron     CAG
                       <2-----[     : 1641]-2>


CAEEL-PGP-4_1a_ 1135  ENITYGLENVSQEQVEKAATLANIHSFVENLPE
                       ENI YGL+  SQEQ+E AA LANIH F+ +LPE
                       ENIMYGLDKCSQEQIEHAARLANIHDFIVSLPE
H_S09_0057179__-1640  gaaatgcgattcgcagcggccgaacgtagaccg
                       aattagtaagcaaataaccgtcataatttgtca
                      gaccgctgtgttaaaaatgattgtttttatttag


CAEEL-PGP-4_1a_ 1168                        GYDTSVGASGGRLSGGQKQRIAIARA
                                             Y+T +GA GG LSGGQKQRIAIARA
                                            QYNTIIGAKGGLLSGGQKQRIAIARA
H_S09_0057179__-1540                        ctaaaaggaggtctggcaccagagcg
                      GTTTGTG  Intron 1   CAGaaacttgcaggttcggaaagtctcgc
                      <0-----[1540 : 1358]-0>actgcaaagttaggatgggtaatttt


CAEEL-PGP-4_1a_ 1194 IVRNPKILLLDEATSALDTESEK                        IVQ
                      IVR+PKILLLDEATSALD ESEK                        +VQ
                      IVRDPKILLLDEATSALDIESEK                        VVQ
H_S09_0057179__-1279 agagcaaccc gggaagcgagaga                        ggc
                      ttgacattttaaccgctatagaaGTATGTT  Intron 2   CAGtta
                      tcatagtgttcacctctcaatgg<0-----[1210 : 1034]-0>gga
                                              1 n [1103 : 1103]


CAEEL-PGP-4_1a_ 1220 EALDKARLGRTCVVIAHRLSTIQ                        NAD
                      EALD+ARLGRTC+VIAHRLSTIQ                        NAD
                      EALDRARLGRTCLVIAHRLSTIQ                        NAD
H_S09_0057179__-1024 ggtgagctgaattgagccttaac                        agg
                      actagcgtggcgtttcagtcctaGTAGGCT  Intron 3   TAGaca
                      atataaagagacatatttatcag<0-----[955  :  739]-0>ctt


CAEEL-PGP-4_1a_ 1246 KIIVCRNGKAIEEGTHQTLLARRGLYYRLVEKQST           1280
                       I+VC +G+  E GTHQ+LL+R+G+YY+ VE+Q+
                      HIVVCCDGRVAEHGTHQSLLSRKGIYYKFVERQNH
```

```
H_S09_0057179__ -729 caggttggcgggcgacctcctcagattatggccac           625
                     atttggaggtcaagcaacttcgagtaaattagaaa
                     tctatttaaacatagcgatgtaaacccgccaaatc


CAEEL-PGP-4_1a_      *
                     *
                     *
H_S09_0057179__ -624 t                                           622
                     g
                     a


H_S09_0057179__ -621 [621  :    1]                                 1
//

Gene 1
EXONS 1640 625
  Exon 1640 1541 phase 0
  Exon 1357 1211 phase 0
  Exon 1033 956 phase 0
  Exon 738 625 phase 0
//

>H_S09_0057179___HAECO_S09_Supercontig_0057179
 1640-625bp_AA
_CAEEL-PGP-4_1a_1135-1280AA_End

ENIMYGLDKCSQEQIEHAARLANIHDFIVSLPEQYNTIIGAKGGLLSGGQKQRI
AIARAIVRDPKILLLDEATSALDIESEKVVQEALDRARLGRTCLVIAHRLSTIQNADHIV
VCCDGRVAEHGTHQSLLSRKGIYYKFVERQNH
//

>H_S09_0057179___HAECO_S09_Supercontig_0057179

ggaaaacatcatgtacggtctggataagtgttctcaagaacaa
atagaacatgcggcacgtcttgcgaatattcatgattttatagttagtcttccagagcaa
tacaatacgatcataggagcaaagggtggtttactgtcgggaggtcagaagcagcgtata
gcaattgctcgtgctattgtcagagatccaaagattctgcttcttgacgaagccaccagt
gcccttgacatagaaagtgagaaggtggtgcaagaagctttagatagagcacgattggga
aggacatgcttagttatagctcatcgtttatctaccatacagaacgctgatcatatcgtt
gtatgttgtgatggacgagtagccgaacatggaacgcaccagtcacttctgtctcgaaaa
ggaatctactacaagttcgtcgaacgacaaaatcac
//
```

A.23 *HAECO*_S09_SUPERCONTIG_0006925 ≈ *CAEEL*-PGP-9

```
genewise
Query protein:        CAEEL-PGP-9___CAEEL-PGP-9_C47A10_1_1-1294AA
Comp Matrix:          blosum62.bla
Gap open:             12
Gap extension:        2
Start/End             local
Target Sequence       H_S09_0006925___HAECO_S09_Supercontig_0006925
Strand:               both
Start/End (protein)   local
Gene Paras:           worm.gf
Codon Table:          codon.table
Subs error:           1e-05
Indel error:          1e-05
Model splice?         model
Model codon bias?     flat
Model intron bias?    tied
Null model            syn
Algorithm             623

H_S09_0006925__     1          Intron      NNN
                               <?-----[     : 2045]-?>
                                 10 n [2036 : 2045]


CAEEL-PGP-9____   205 LTLIMMSLSPFMMICGLFLAK                          LLATA
                      LTLIMMSL+PFM+ICG F+AK                          L+ATA
                      LTLIMMSLAPFMIICGAFIAK                          LMATA
H_S09_0006925__  2046 cacaaatcgctaaatggtaga                          cagag
                      tcttttctccttttggcttcaGTTTGTC  Intron 1   CAGttccc
                      cggtggggagtgcccagcttg<0-----[2109 : 2174]-0>ggtaa


CAEEL-PGP-9____   231 ATKEAKQYAVAGGIAEEVLTSIRTVIAFNGQEYECK
                      AT+EAK+YAVAGGIAEEVLTSIRTVIAFNGQ YEC+
                      ATREAKKYAVAGGIAEEVLTSIRTVIAFNGQPYECE
H_S09_0006925__  2190 gacggaatgggggagggcataaagagtagcctgtg
                      ccgacaaactcggtcaattcctgcttctagacaaga
                      gccgcggttgctactaggcaacgcccactcactgca


CAEEL-PGP-9____   267                R          YEDALEHGKKTGIKKSFLIGAGL
                                     R          Y+ ALE GK TGIKKSF IG GL
                                     R          YQKALEDGKSTGIKKSFYIGVGL
H_S09_0006925__  2298 a          [agg]       tcagtgggatagaaatttagggc
                      gGTATGAT  Intron 2   CAG aaactaagaccgtaactatgtgt
                       <2-----[2300 : 2993]-2>gcaagaactaagacgatttttcttg


CAEEL-PGP-9____   291 ASFFVIIYASYCLAFWVGTNFVYSGRLESGTVLT          324
                        F+I+++SYCLAFWVGT+FV+ G++  GTV+T
                      GITFLIMFSSYCLAFWVGTDFVFKGQMNGGTVMT
H_S09_0006925__  3064 gaatcaatttttcgttggaqtgtagcaaggagaa         3165
                      gtctttttccagtctgtgcatttagataggcttc
                      aagcgcgccattgatgccgttccatagtaatgga


Alternative splice

CAEEL-PGP-9____   325                          VFFSVMMGSMALGQAGQQFATIGTAL
                                               VFFSVMMGSMALGQAG QFA +GTA+
                                               VFFSVMMGSMALGQAGPQFAVLGTAM
H_S09_0006925__  3166                          gtttgaagtagcgcggcctggcgaga
                      GTACGTT  Intron 3   AAGtttctttgctctgacgcatcttgcct
                      <0-----[3166 : 3217]-0>tcccaggttggcagtgtatttctgtg


CAEEL-PGP-9____   351 GAAASLYEVIDR                          IPEIDAYSTEGQTP
                      GAA SLY++IDR                           PEID+YST+G  P
                      GAAGSLYQIIDR                          EPEIDSYSTDGVKP
H_S09_0006925__  3296 ggggtctcaagc                          gcgagtttagggac
                      gccgctaattagGTATGTT  Intron 4   AAGacatacaccagtac
                      tatatctgtctg<0-----[3332 : 3460]-0>aaatcgcaattgaa


CAEEL-PGP-9____   377 SKISGRISVNKVEFTYPTRADVKILK
                      S + G+++V+ ++FTYPTR DV ILK
                      SNLKGKVTVSNLKFTYPTRPDVPILK
H_S09_0006925__  3503 tacagagagtacatatcaccggcaca
                      catagatctcatatcaccgcatcttaGTGTGCC  Intron 5   CAG
                      gtcaggaatctgacctaaaatttttg<0-----[3581 : 3634]-0>


CAEEL-PGP-9____   403 GVSLDAQPGQTVALVGSSGCGKSTIIQLLQRFYNPDAGQ          441
                      GVS +A PG+T+ALVGSSGCGKSTIIQLL R+YNP+ G+
                      GVSFEANPGETIALVGSSGCGKSTIIQLLLRYYNPEDGK
H_S09_0006925__  3635 ggttggacggaagtggtagtgaaaaacttccttacggga         3751
                      gtctacacgactcttgcgggggagcttatttgaaacaaga
                      aaacaatttagaaggtctttaatcacggagatcttgtag


H_S09_0006925__  3752 GTATGCC  Intron 6
                      <0-----[3752 : 4965]-0>
                        10 n [4061 : 4070]

Alternative splice

CAEEL-PGP-9____   325 V             -          FFSVMMGSMALGQAGQQFATIGT
                      V                        FFSVMMGSMALGQAG QFA +GT
                      V             R          FFSVMMGSMALGQAGPQFAVLGT
H_S09_0006925__  3166 gc          [cgt]        ttgaagtagcgcggcctggcga
                      t GTTTCTT  Intron 3   AAG tctttgctctgacgcatcttgc
                      a <1-----[3170 : 4146]-1>tccaggtcggcaatatgttcctg
                         10 n [4061 : 4070]


CAEEL-PGP-9____   349 ALGAAASLYEVIDRI              P          EIDAYSTE
                      A+GAA SLY++IDR+                         ID+YST+
                      AMGAAGSLYQIIDRV              E          -IDSYSTD
H_S09_0006925__  4218 gaggggtctcaagcgg           [gaa]        agtttag
                      ctgccgctaattagt GTTCAAA  Intron 4   CAGa tacacca
                      tgtacatctgtttga <1-----[4264 : 4382]-1>a ctgtagt


CAEEL-PGP-9____   373 GQTPSKISGRISVNKVEFTYPTRADVKILK
                      G  PS + G+++V+ ++FTYPTR DV ILK
                      GVKPSNLKGKVTVSNLKFTYPTRPDVPILK
H_S09_0006925__  4406 ggactacagagaqtacatatcaccggcaca
                      gtaccatagatctcatatcaccgcatcttaGTTTGCC  Intron 5
                      cgaagtcagaaacatgacctagaatttttg<0-----[4496 : 4552]


CAEEL-PGP-9____   403    GVSLDAQPGQTVALVGSSGCGKSTIIQLLQRFYNPDAGQ      441
                         GVS +A PG+T+ALVGSSGCGKSTIIQLL R+YNP+ G+
                         GVSFEANPGETIALVGSSGCGKSTIIQLLLRYYNPEDGK
H_S09_0006925__  4550    ggttggacggaagtggtagtgaaaaacttccttacggga      4669
                      CAGgtctacacgactcttgcgggggagcttatttgaaacaaga
                      -0>acacaatttagaaggtctttaatcacggagatcttgtgg


H_S09_0006925__  4670 GTATGCC  Intron 6
                      <0-----[4670 : 4965]-0>


H_S09_0006925__          Intron 6   CAG
                      <0-----[     : 4965]-0>


CAEEL-PGP-9____   442 ILIDDIPIEDFNIKYLRQLVGVVSQE
                      I  D + I+ +NI++LR  VGVVSQE
                      ITRDGVEIDKINIEFLRNYVGVVSQE
H_S09_0006925__  4966 aaaggggagaaaagtccatggggtcg
                      tcgagtataatatattgaatgttcaa
                      tgattgaccattcactacctacctag


CAEEL-PGP-9____   468 PNLFNTSIEQ                          NIRYGRSDVSDEDIAR
                      P LFNT+IEQ                          NIRYGR +V+D +I
                      PMLFNTTIEQ                          NIRYGRENVTDAEITA
H_S09_0006925__  5044 cactaaaagc                          aactgcgagagggaag
                      ctttacctaaGTTGGTC  Intron 7   CAGatgaggaatcacatcc
                      ggtctggtag<0-----[5074 : 5124]-0>ccccaagtcctgacct


CAEEL-PGP-9____   494 ALKEANAADFIKTFPE                          GLNTLVGDRG
                      AL++ANA +F+++FP+                          G+ T VGDRG
                      ALRKANAYNFVQSFPD                          GIYTNVGDRG
H_S09_0006925__  5173 gccagagtatgcttcg                          gataagggcg
                      ctgacacaattactcaGTAAGTG  Intron 8   CAGgtacatgagg
                      tataactcctgagtttt<0-----[5221 : 5281]-0>gccgcctctc


CAEEL-PGP-9____   520 VQMSGGQKQRIAIARALVRNPKILLLDEATSALDAESESIVQSALEN
                      QMSGGQKQRIAIARALVR+PKILLLDEATSALDAESE IVQ ALEN
                      TQMSGGQKQRIAIARALVRDPKILLLDEATSALDAESEHIVQQALEN
H_S09_0006925__  5312 acatggcaccagagcgcgagcaacccgggaagcgggagcagccgcga
                      catcggaaagtctcgcttgacattttaaccgctacagaattaactaa
                      tggcccgggtactttggcatgatcggtgctctcccatacctaggggc


CAEEL-PGP-9____   567                          ASRGRTTIVIAHRLSTVRNADKIIVM
                                               AS+GRTTIV+AHRLST+RNADKI+ M
                                               ASKGRTTIVVAHRLSTIRNADKIVAM
H_S09_0006925__  5453                          gtagaaaagggccctaacaggaagga
                      GTAAGTT  Intron 9   TAGccaggcctttcagtcctgacaatttct
                      <0-----[5453 : 5505]-0>gcaaaaccctttaggatattcgttgg


CAEEL-PGP-9____   593 KAGQ                          VMEVGTHETLIEQKGLYHELVH
                      K G                           VMEVGTH+ LI +KGLYHELV+
                      KNGE                          VMEVGTHDELIARKGLYHELVN
H_S09_0006925__  5584 aagg                          gagggacggcagctcgcga
                      aagaGTGCTGT  Intron 10  CAGttatgcaaattcgagtaaatta
                      gtag<0-----[5596 : 5656]-0>tgaccattggcacgacctggct


CAEEL-PGP-9____   619 AQVFADVDD              K          PK          630
                      AQVFADVDD                          P
                      AQVFADVDD              T          PD
H_S09_0006925__  5722 gcgtggggga          [act]        cg          6159
                      cattcataa GTAAGTT  Intron 11  CAGcca
                      cggcccgct <1-----[5750 : 6151]-1>ttc


CAEEL-PGP-9____   631 K                                              631

                      !!!!!
H_S09_0006925__  6160 cttac                                          6164


CAEEL-PGP-9____   632 KEAERRMSRQTSQRKGSVNFKTQESQVDEKPGAPPAPEAAEKEIKRLKK
                        + R  SQ         T+ +      GA   P  AEK+++RLKK
                      SVVISAFQRLKSQ------MSTENAAGG---GAQNDPVKAEKDLERLKK
H_S09_0006925__  6165 tggatgtcccatc      atagagggg   ggcagcgaggagtgccaa
                      ctttcctagtaca      tccaaccgg   gcaaactacaaatagtaa
                      cgctcttggcgta      gacatttta   ttgctagaggacgaagga


CAEEL-PGP-9____   681                E          LEEEGAVKANLFKILRYARPEWI
                                     E          LEEEGA KANLFKIL YARPEW
                                     E          LEEEGAAKANLFKILGYARPEWP
H_S09_0006925__  6285 g          [gaa]        cgggggagactaacgtgcccgtc
                       GTGATTC  Intron 12  CAGataaagccacattattgacgcagc
                       <1-----[6286 : 6347]-1>acagatccagtccactcttatagt


CAEEL-PGP-9____   705 YIFFAIIAALIQGAVMPAFSLFFSQIIN
                      +I +A+ ++++QG V+PAFSLFFSQII+
                      FIALAVTSSIVQGCVFPAFSLFFSQIID
H_S09_0006925__  6419 tagcggattagcgtgtcgttctttcaag
                      ttctctcccttaggttcctctttcattaGTACGCG  Intron 13
                      tccccctgctacttcaacactcaatcc<0-----[6503 : 6556]


CAEEL-PGP-9____   733    VFS-NPDRDQMKKDGHFWALMFLVLAAVQGTSMLFQ
                         VFS  P    +K DGHFWALMFLVL    Q  +ML+Q
                         VFSKQPGDPTLKSDGHFWALMFLVLGGTQAMTMLIQ
H_S09_0006925__  6554    gttaccggcacaaggcttgcatcgcggacgaaacac
                      CAGtcaacgactagagatgctttttggcactcttta
                      -0>atcgaattggcgtctttggtgccggtctgggcgctg


CAEEL-PGP-9____   768                          CSLFGVAAERLTMRIRSKVYRNVLRQ
                                               C  FG++AERLTMR+RSK+++NV+R
                                               CFFFGLSAERLTMRLRSKIFQNVMRM
H_S09_0006925__  6665                          ttttgttggccaactctaatcagaaa
                      GTATACC  Intron 14  CAGgtttgtccagtctgtgcattaattgt
                      <0-----[6665 : 6717]-0>cccctgacagcagaaaaatcatggag


CAEEL-PGP-9____   794 DATYFDMPKHSPGRITTRLATDAPNIKS
                      DATYFDMP+HS G+ITTRLATDAPN+KS
                      DATYFDMPRHSAGKITTRLATDAPNVKS
H_S09_0006925__  6796 ggattgaccctggaaaactgaggcagat
                      accatatcgaccgatccgtccaccatacGTACGGT  Intron 15
                      tcccccgtctataactccggataatgag<0-----[6880 : 7512]


CAEEL-PGP-9____   822    AIDYRLGSIFNAIASVGGGLGIAFYYGWQMAFL
                         A+DYR GS+F+++ S+  G+GIAFY+GWQMA+L
                         ALDYRFGSVFSSVVSICCGIGIAFYFGWQMALL
H_S09_0006925__  7510    gcgtctgtgtatggtattgagagtttgtcagct
                      CAGctaagtgcttgcttctgggtgtctatggatctt
                      -0>ttcctctgcctactcactcatcacttagggata


CAEEL-PGP-9____   855                          VMAIFPFMAVGQALMMKYHGGSATSD
                                               +AIFP+ AVGQA++M++  G AT+D
                                               TIAIFPLAAVGQAIQMRFMSGRATAD
H_S09_0006925__  7612                          aagatctgggcgacaatatgcgagg
                      GTGAGTT  Intron 16  CAGctcttctcctgactatgttcggccca
                      <0-----[7612 : 7671]-0>acctctgttggattggacgtgtcact


CAEEL-PGP-9____   881 AKEMENAGK                          TAMEAIENIRTVQALTL
                      AKEMEN+GK                           AMEAIENIRTVQALTL
                      AKEMENSGK                          IAMEAIENIRTVQALTL
H_S09_0006925__  7750 gagagaaga                          agaggagaacagcgtat
                      caataaggaGTGGGTC  Intron 17  CAGtctactaatgctactct
                      ggggactag<0-----[7777 : 7829]-0>taggccgttagggaggg


CAEEL-PGP-9____   907 QTKLYNIFCSHLDAPHGGNISKAIIR
                      + +L+  FC HLD PH  +  KA+I+
                      ERRLHAQFCHHLDGPHKTSRRKALIQ
H_S09_0006925__  7881 gccccgcttcccggccaaaacagcac
                      aggtacatgaatagcaacgggacttaGTGGGGA  Intron 18  AAG
                      gtgttcgctttacagtaccaggtttg<0-----[7959 : 8038]-0>


CAEEL-PGP-9____   933 GLTYGFANSIQFFTYAAAFRFGLFLIFDKNVLMEPENVL
                      G++YGFA+SI +F YA+ FRFGL+LI   N  +   NVL
                      GVSYGFASSIFYFLYASCFRFGLWLIV--NGTIHSMNVL
H_S09_0006925__  8039 ggttgtgaaatttctgtttctgttcag  agaactaagc
                      gtcagtcggttattaccgtgtgtgttt  agctactatt
                      tcatacccctcccgcaatctcgggtcc  ccctcggttc


CAEEL-PGP-9____   972                R          VLFAISFSFGTIGFAASYFPEYI
                                     R          VLFAISF+ G++GFA+SYFPEYI
```

```
                                    R          VLFAISFTAGSMGFASSYFPEYI
H_S09_0006925__ 8150 a           [agg]        gctgattaggaagtgatttcgta
                     gGTGAGCC  Intron 19  CAG tttctctccggtgtcgcatcaat
                      <2-----[8152 : 8272]-2>gaatgttcttatgatctgtccgcc


CAEEL-PGP-9____  996 KATFAAGLIFNMLEEEPRIDGMTSSGTYPQLSGEVKLNKVFFRYPERPA
                     KATFAAG+IF+MLEEEPRIDGMT++G  P+++G VKLNKV+F+YPERP
                     KATFAAGIIFHMLEEEPRIDGMTNNGKKPKITGAVKLNKVYFKYPERPD
H_S09_0006925__ 8343 agatgggaatcatgggccaggaaaagaacaaagggacaagttatcgccg
                     acctccgtttattaaacgtagtcaagaacatcgctataatataacagca
                     aggttctccccggaaaaacccggtccgacgtctccggtactcatcagac


CAEEL-PGP-9____ 1045 VPILQGLNVH                          VKPGQTLALVGPSGCG
                     VPILQGL+++                          VKPG+TLALVGPSGCG
                     VPILQGLDIN                          VKPGETLALVGPSGCG
H_S09_0006925__ 8490 gcacgcgaa                           gacggacgcggcagtg
                     tcttagtataGTAAGCT  Intron 20  CAGtacgactcttgcgggg
                     tcataacccc<0-----[8520 : 8608]-0>gaccgttcgtgcccca


CAEEL-PGP-9____ 1071 KSTVISLLERLYDPLEGAV                        TVDNNDL
                     KSTVISLLERLYD L+G+V                         +D NDL
                     KSTVISLLERLYDALDGSV                        EIDGNDL
H_S09_0006925__ 8657 atagatccgactggtggtg                        gaggagc
                     accttcttagtaactagctGTGAGTA  Intron 21  TAGatagaat
                     gcgtatacagccttgcttt<0-----[8714 : 9178]-0>acttcta


CAEEL-PGP-9____ 1097 RQMNPKHLRKHIALVSQEPILFDTSIRENIVYGLQPGEYTHEQIETACS
                     R++NP HLR HIALVSQEPILFD SIR+NI+YGL PG  +   +
                     REVNPTHLRAHIALVSQEPILFDRSIRDNILYGLPPGSVSDAAVHEVAQ
H_S09_0006925__ 9200 cggacacccgcagtgtcgcactgatacgaactgcccgtgaggggcgggc
                     gataccatgcatcttcaactttagctgaattagtccgctgacctaatca
                     tagcattgacctcgagggactccacctctccctattccttgtgcacaa


CAEEL-PGP-9____ 1146 KANIHKFIDELPD                           GYETRVGEKGTQL
                     +ANIHKFI +LPD                           GY TR GEKGTQL
                     RANIHKFIMDLPD                           GYNTRAGEKGTQL
H_S09_0006925__ 9347 cgaacataagccg                           gtaacgggagact
                     gcataatttatcaGTAAGTT  Intron 22  CAGgaacgcgaagcat
                     tccccaccgcgat<0-----[9386 : 9456]-0>tcctcaaaaatag


CAEEL-PGP-9____ 1172 SGGQKQRIAIARALIRNPKILLLDEATSALDTESEK
                     SGGQKQRIAIARALIRNPKILLLDEATSALDTESEK
                     SGGQKQRIAIARALIRNPKILLLDEATSALDTESEK
H_S09_0006925__ 9496 tggcaccagagcgcaaacaacccgggaagtgagaga
                     cggaaagtctcgcttgacatttaaccgctacagaa
                     tacaaaacccattccatgatggtcgtcttgtcacgg


CAEEL-PGP-9____ 1208                          QVQVALDAAAKDRTCIVVAHRLSTIV
                                               VQ ALD A++ RTCIVVAHRLST+V
                                              VVQEALDKASEGRTCIVVAHRLSTVV
H_S09_0006925__ 9604                          ggcggcgagtggcataggggccctagg
                     GTTCTAT  Intron 23  CAGttaactaaccaggcgtttcagtcctt
                     <0-----[9604 : 9681]-0>tgagcccgaaaccctctgatgggttc


CAEEL-PGP-9____ 1234 NAGCIMVVKNGQVVEQ              G          THNELIA
                     NA CIMVVK G+VVE+               G          TH+EL+
                     NANCIMVVKGGKVVEK               G          THSELMQ
H_S09_0006925__ 9760 agataaggaggagggag           [gga]        acagcac
                     acagttttaggattaa GTACAGC  Intron 24  CAGgcagatta
                     tcttagacgaaataaa <1-----[9809 : 9868]-1>acttagga


CAEEL-PGP-9____ 1258 KRGAYFALTQKQS                                  1270
                      +GAY+ALTQKQ+
                     AKGAYWALTQKQT
H_S09_0006925__ 9892 gaggttgcacaca                                  9930
                     caqcagctcaaac
                     cgtacggtaggaa


CAEEL-PGP-9____ 1271 SNQSGGAFDTSEALDDDDDDHVKF                       1294

                     LAKG----------------------
H_S09_0006925__ 9931 tgag                                           9942
                     tcag
                     gcga


CAEEL-PGP-9____      *
                     *
                     *
H_S09_0006925__ 9943 t                                              9945
                     g
                     a

H_S09_0006925__ 9946      [9946 :16712]                            16712
                      16 n [11408:11423]
                      10 n [15181:15190]

//

Gene 1
EXONS 2046 9942
  Exon 2046 2108 phase 0
  Exon 2175 2299 phase 0
  Exon 2994 3165 phase 2
  Exon 3217 3331 phase 0
  Exon 3461 3580 phase 0
  Exon 3635 3751 phase 0

  Exon 2994 3169 phase 2
  Exon 4147 4263 phase 1
  Exon 4383 4495 phase 1
  Exon 4553 4669 phase 0

  Exon 4966 5073 phase 0
  Exon 5125 5220 phase 0
  Exon 5282 5452 phase 0
  Exon 5506 5595 phase 0
  Exon 5657 5749 phase 0
  Exon 6152 6159
       6160 6164
       6165 6285 phase 1
  Exon 6348 6502 phase 1
  Exon 6557 6664 phase 0
  Exon 6718 6879 phase 0
  Exon 7513 7611 phase 0
  Exon 7672 7776 phase 0
  Exon 7830 7958 phase 0
  Exon 8039 8151 phase 0
  Exon 8273 8519 phase 2
  Exon 8609 8713 phase 0
  Exon 9179 9385 phase 0
  Exon 9457 9603 phase 0
  Exon 9682 9808 phase 0
  Exon 9869 9942 phase 1
//

Making a R in phase 2 intron

Making a R in phase 1 intron
Making a E in phase 1 intron

Making a T in phase 1 intron
Making a E in phase 1 intron
Making a R in phase 2 intron
Making a G in phase 1 intron


>H_S09_0006925___HAECO_S09_Supercontig_0006925
```

```
_2046-3165bp_AA
_3166-3751bp_AA_Alternative_splice
_3166-4669bp_AA_Alternative_splice
_4966-6159bp_AA
_6160-6164bp_AA_Frameshift
_6165-9942bp_AA
_CAEEL-PGP-9_205-324AA_325-441AA_Alternative_splices_442-
630AA_631AA_Frameshift_632-1294AA_End

LTLIMMSLAPFMIICGAFIAKLMATAATREAKKYAVAGGIAEEVLTSIRTVIAFNGQPYE
CERYQKALEDGKSTGIKKSFYIGVGLGITFLIMFSSYCLAFWVGTDFVFKGQMNGGTVMT

VFFSVMMGSMALGQAGPQFAVLGTAM
GAAGSLYQIIDR
EPEIDSYSTDGVKP
SNLKGKVTVSNLKFTYPTRPDVPILK
GVSFEANPGETIALVGSSGCGKSTIIQLLLRYYNPEDGK

VRFFSVMMGSMALGQAGPQFAVLGTAMGAAGSLYQIIDRVEIDSYSTDGVKPSNLKGKVT
VSNLKFTYPTRPDVPILK
GVSFEANPGETIALVGSSGCGKS
TIIQLLLRYYNPEDGK

ITRDGVEIDKINIEFLRNYVGVVSQEPMLFNTTIEQNIRYGREN
VTDAEITAALRKANAYNFVQSFPDGIYTNVGDRGTQMSGGQKQRIAIARALVRDPKILLL
DEATSALDAESEHIVQQALENASKGRTTIVVAHRLSTIRNADKIVAM
KNGE
VMEVGTHDELIARKGLYHELVNAQVFADVDDTPD

SVVISAFQRLKSQM
STENAAGGGAQNDPVKAEKDLERLKKELEEEGAAKANLFKILGYARPEWPFIALAVTSSI
VQGCVFPAFSLFFSQIID
VFSKQPGDPTLKSDGHFWALMFLV
LGGTQAMTMLIQCFFFGLSAERLTMRLRSKIFQNVMRMDATYFDMPRHSAGKITTRLATD
APNVKSALDYRFGSVFSSVVSICCGIGIAFYFGWQMALLTIAIFPLLAAVGQAIQMRFMSG
RATADAKEMENSGKIAMEAIENIRTVQALTLERRLHAQFCHHLDGPHKTSRRKALIQGVS
YGFSSIFYFLYASCFRFGLWLIVNGTIHSMNVLRVLFAISFTAGSMGFASSYFPEYIKA
TFAAGIIFHMLEEEPRIDGMTNNGKKPKITGAVKLNKVYFKYPERPDVPILQGLDINVKP
GETLALVGPSGCGKSTVISLLERLYDALDGSVEIDGNDLREVNPTHLRAHIALVSQEPIL
FDRSIRDNILYGLPPGSVSDAAVHEVAQRANIHKFIMDLPDGYNTRAGEKGTQLSGGQKQ
RIAIARALIRNPKILLLDEATSALDTESEKVVQEALDKASEGRTCIVVAHRLSTVVNANC
IMVVKGGKVVEKGTHSELMQAKGAYWALTQKQT
LAKG
//

>H_S09_0006925___HAECO_S09_Supercontig_0006925
_2046-3165bp
_3166-3751bp
_3166-4669bp
_4966-6159bp
_6160-6164bp
_6165-9942bp

ctcacgctgattatgatgtcgctggcaccgtttatgatcatctgcggagcgttcattgct
aagctgatggctacagcagcgacccgcgaggccaagaagtatgctgtggccggtggaatc
gctgaagaggtgctcacatcaatcaggaccgtcatcgcattcaatggccaaccctatgag
tgcgaaaggtaccaaaaagcgttagaagacggtaaatcaacgggaatcaagaaatctttt
tatattggcgttggtctgggaataacgttcctgatcatgttctcctcatattgctctggca
ttttgggtcggcacggattttgtcttcaaaggtcaaatgaatggaggaactgtgatgaca

gttttcttctccgtaatgatggg
ttctatggcgctcggacaggctgggcctcaatttgctgttctcggtacggctatgggtgc
agctggatctctctatcagattatcgatcgg
gaaccagaaattgactcgta
ctcaacagatggtgtgaaaccatcgaatctcaaagggaaggtaacagtttccaatctgaa
attcacctatccaacacgaccagatgttcctattcttaag
ggagtatcattcgaagcaaatcctgg
tgaaacgatagcattggtgggttccagtggttgtggaaaaagtaccataatccagttgtt
actgcgatattacaatcctgaggatggaaag

gtac
gt
ttcttctccgtaatgatgggttccatggcgctcggacaagctggacctcagttt
gctgtcctcggtacggctatgggtgcagccggatctctctatcagattattgatcgggta
gaaatcgattcgtattcaacggatggcgtgaaaccatcgaatctcaaagggaaagtaaca
gtctcaaatctgaaattcacctatccaacgcgaccagatgttcctattcttaag
ggagtctca
ttcgaagcaaatcctggtgaaacgatagcattggtgggttccagtggttgtggaaaaagt
accataatccagttgttactgcgatattacaatcctgaggatgggaag

attacgagagat
ggtgtggaaatcgacaaaattaatatcgaattccttcgaaactacgttggagtcgtctct
caagagccgatgcttttcaatacgacgattgaacagaacatccgctacggacgagagaat
gtcaccgatgcggaaatcaccgctgctctacgtaaagcaaacgcttacaactttgtgcaa
tcgtttcctgatgggatctacacgaacgtcggtgaccgtggcactcagatgtccggcggc
cagaagcagcgtatagccattgctcgtgcgctggtcagagatccgaaaattctcctgctg
gatgaggccactagcgctctcgacgccgaaagtgaacacatcgttcaacaggcgctggag
aacgcgtccaaaggaagaacaaccatcgtcgttgctcatcgactgtcgacaattcgaaat
gctgacaagattgttgcgatgaagaatggagag
gttatggaagtcggcacacatgatgag
ctgatcgcacgcaagggactctaccatgagctggtcaatgcccaggtgttcgccgacgtg
gacgatactcctgac
cttac

tccgtggtcatttccgcttttcagcggctcaagtctcaaatgtcaaccgaaaatgctgct
ggtggaggtgctcagaacgatccagtgaaagcggagaaagacttggaacgactgaagaaa
gaactcgaagaggaaggtgccgccaaagcgaatctcttcaaaatccttggctatgctcga
cctgaatggccttttatcgccctcgccgtcacctcttcgatcgttcaaggctgttttttc
ccagcattctcactcttttctcacaaattatcgac
gtatttttcaagcaaccaggtgatccgacg
ctcaagagtgacggtcatttgggcgcttatgttcctcgtgctgggtggcgcactcaggcg
atgaccatgctcattcagtgcttcttcttcggtttgtcagccgaacggctcacaatgcga
ttacgatcaaaaatttttcaaaatgtgatgagaatggatgccacctacttcgacatgcct
cgccattcagctggaaaaatcactacccgcttggcgacagatgcaccaaatgtgaaatcg
gctcttgactaccgtttcggttcggtcttcagttcagtcgtttccatatgctgtggcata
ggtatcgcattctatttttggatggcagatggcacttttaacaatcgccatttttccctttg
gctgctgtggggcaagctattcagatgagattcatgtctgggcgtgccacagccgatgcg
aaggagatggaaaacagtggaaagattgcaatggaggccatcgagaatattcgaacggtg
caggcattgacgttggagcgtcggcttcatgccccagttctgtcatcatctagacggaccg
cataaaaccagcagacggaaggctcttattcagggtgtctcatatggattcgccagcagt
atcttctactcctgtacgcatcatgtttccgtttcgggttgtggcttatcgtcaacggc
accattcactcgatgaatgttctcagggtactatttgcgatttctttcactgctggaagt
atgggatttgccagttcgtatttccccgagtacatcaaagcgacgtttgctgccggtatc
atcttccacatgttggaagaagaaccacgaatcgacggcatgacgaataacggcaagaaa
cccaagattaccggtgccgtcaagctgaataaagtctatttcaaatatcccgaacggcca
gacgttcccatacttcaaggactcgacatcaacgtgaaacccggcgagacttctgccctg
gttgggcccagcggctgcggaaagtccacggttatatctactcgaaaggctctacgat
gctttggacggttctgttgaaatcgatggtaacgatctacgtgaagtgaacccaactcat
ctgcgagcccacattgccttggtatcgcaggagccaatccttttcgacagatccatccgt
gacaatatcctctacggccttccacctggttccgtcagtgatgcggctgtgcacgaagtc
gcacaacgtgccaacaccccaaaattcatcatggacctgccagatggttacaacactcgc
gcaggaaaaggaggaactcaattgtctggaggccaaaacaacgaatcgccatcgcacgt
gctctcatcagaaatccgaaaattctgctgcttgacgaggctaccagtgctttggatacc
gaaagcgagaaggttgtgcaagaggccctcgacaaggccatcagaaggccgcacctgtatc
gttgtggcacatcggctgtcgactgttgtcaatgccaattgtataatggtagtcaaggga
ggaaaagttgtagaaaaaggaacccatagtgaactgatgcaagccaagggtgcatactgg
gcgcttacacagaagcaaaca
ttggccaaggga
//

genewise
Query protein:       CAEEL-PGP-9____CAEEL-PGP-9_C47A10_1_1-1294AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            global
Target Sequence      H_S09_0006925___HAECO_S09_Supercontig_0006925
                                    _10000-27070bp
Strand:              both
Start/End (protein)  global
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
```

```
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623

H_S09_0006925__ 9946      [9946 :16712]
                          16 n [11408:11423]
                          10 n [15181:15190]


CAEEL-PGP-9____     1 MGLFKKK-KDDSSSEGSSEKKEE              -
                      MG +KKK K   S E  ++ +EE
                      MGFLKKKGKAADSKESENDNEEE              K
H_S09_0006925__ 16713 agttaaagagggaagtgagaggga           [aag]
                      tgttaaagaccagaacaaaaaaaaGTGGGTT  Intron 1   CAG
                      gtcaggagaactcagagtccaaa <2-----[16784:18124]-2>g
                                              328 n [17038:17365]


CAEEL-PGP-9____    23 APPPPKISIFQL                        YRYTSTVDRLMLAV
                         PK SI QL                         +RYT+T D+L+L +
                      KDDAPKASIPQL                        FRYTTTFDKLLLLI
H_S09_0006925__ 18126 agggcagaaccc                        tctaaatgattcta
                      aaacccacgtcatGTTCGTT  Intron 2   CAGtgaccctaattttt
                      attagactcaag<0-----[18162:18350]-0>ctcacaccgggcgc


CAEEL-PGP-9____    49 GIIVSCATGVGLPLMSIIMGNVSQNF              V
                      G  V+  TG+GLP+M+IIM  V   F
                      GSTVAMGTGMGLPMMAIIM--VRHEF              T
H_S09_0006925__ 18393 gtaggagagagccaagaaa  gacgta          [act]
                      gcctctgcgtgtcttcttt  tgaat GTGCGGA  Intron 3
                      tatctgtttgacaggttcg  agtac <1-----[18466:18517]


CAEEL-PGP-9____    75      TLGTIFLDPNSTASEKAA              A
                           T+ T   LD +ST +
                           TI-TYGLDASSTTAPYRL              T
H_S09_0006925__ 18515      aa atgtggaaaagctaca           [act]
                      CAGcct cagtacggccccagt GTCTCCC  Intron 4   CAGc
                      -1>tca ccagcatcccctacgt <1-----[18572:19513]-1>t


CAEEL-PGP-9____    95 RAEFSHEVIQNCLKYVYLGCGIFAAGFLQ
                        +F H+V QNCLKYVYLGCGIFAA  +Q
                      FQQFEHDVTQNCLKYVYLGCGIFAAATIQ
H_S09_0006925__ 19516 tcctgcggacattatgtcgtgatgggaac
                      taataaatcaagtaatatgggtcccctaGTATGGT  Intron 5
                      taataccgaactgacccttaacaaagtg<0-----[19603:20636]
                                                   13 n [19830:19842]


CAEEL-PGP-9____   124    ASCFMVICEKLSNRFRRQFFHSVMRQEIAWYDKNTSGTLSNKLF
                         A+CF+ + E L N++RRQFF S++RQ+I W+DKN SGTL+ KLF
                         ATCFLTVGENLVNQLRRQFFKSILRQDIPWFDKNGSGTLATKLF
                         gattcagggacgactaccttatacccgacttgaagtgatgaact
H_S09_0006925__ 20634 AAGccgttctgaattaatggattacttgaatcgtaaagcgctccatt
                      -0>gatcgtcaacggtaggagtcgattcactagttgttaaagcaggc


CAEEL-PGP-9____   168              D          NLERVREGTGDKVGLAFQMMAQF
                                   D          NLERV+EGTGDKVGL +Q +AQF
                                   D          NLERVKEGTGDKVGLMIQYVAQF
H_S09_0006925__ 20769 g          [gac]        acgcgaggaggaggcaactggct
                      aGTGAGTG  Intron 6   AAG atagtaagcgaatgtttaatcat
                       <2-----[20771:20829]-2>ctcaacgaacccactggcgtgagc


CAEEL-PGP-9____   192 IGGFAVAFTYDWLLTLIMMSLSPFMMICGLFLAK
                      +GGF VAFTYDW LTLIMMSL+PFM++CG F+A+
                      FGGFIVAFTYDWKLTLIMMSLAPFMIMCGAFIAR
H_S09_0006925__ 20900 tggtaggtatgtacacaaatcgctaaatggtaga
                      tggtttctcaagatcttttctccttttggcttcg
                      ctcccgacctcgacggtggcgtacgcgcaatttg


CAEEL-PGP-9____   226                        LLATAATKEAKQYAVAGGIAEEVLTSI
                                             L+ATAAT+EAK+YAVAGGIAEEVLTS+
                                             LMATAATREAKKYAVAGGIAEEVLTSM
H_S09_0006925__ 21002                        tagaggacggaatgggggagggcata
                      GTAAGCT  Intron 7   CAGttcccccgcaaactcggtcaattcct
                      <0-----[21002:21109]-0>ggctagttgcagtagtaaacaggttgg


CAEEL-PGP-9____   253 RTVIAFNGQEYECK              R          YEDALEHGKKT
                      RTV+AFNGQ YEC+              R          Y+ ALE GK T
                      RTVMAFNGQPYECE              R          YKQALEDGKST
H_S09_0006925__ 21191 aagagtagcctgtga           [agg]        tacgcgggata
                      gcttctagacaagagGTACCCT  Intron 8   CAG aaactaagacc
                      gttgtccagttgca <2-----[21235:21672]-2>gcggagattata


CAEEL-PGP-9____   279 GIKKSFLIGAGLASPFVIIYASYCLAFWVGTNFVYSGRLESGTVL  323
                      GIKKS+ +G G A  F+I+++SYCLAFWVGT+FVY+ R++ GTV+
                      GIKKSLYVGIGFAITFLILFSSYCLAFWVGTDFVYNNRMQGGTVM
H_S09_0006925__ 21707 gaaatttggagtgaatcatttttttgttggagtgtaacacggaga21841
                      gtaactatgtgtctctttttccagtctgtgcattaaagtaggctt
                      aagacgttttccatatgcgctctcgatgccgtccctcggatatgg


CAEEL-PGP-9____   324 T-                                              324
                      T
                      TA
H_S09_0006925__ 21842 ag
                      cc
                      ga


                      !
H_S09_0006925__ 21848 a                                              21848


CAEEL-PGP-9____   325                        VFFSVMMGSMALGQAGQQFATI
                                             VFF+VMMGSMALGQAG QFA +
                                             VFFAVMMGSMALGQAGPQFAVL
H_S09_0006925__ 21849                        gttggaagtagcgcggcctggc
                      GTTTTTT  Intron 9   TAGtttctttgctctgacgcatctt
                      <0-----[21849:21910]-0>cttctggctggtaatatgttat


CAEEL-PGP-9____   347 GTALGAAASLYEVIDR                        IPEIDAYSTE
                      GTA+GAA SLY++IDR                         PEID+YS E
                      GTAMGAAGSLYQIIDR                        EPEIDSYSQE
H_S09_0006925__ 21977 gagaggggtctcaaga                        gcgagtttcg
                      gcctgccgctaattagGTAAGTC  Intron 10  CAGacatacacaa
                      tatgtacatctgcctg<0-----[22025:22129]-0>aagccccaga


CAEEL-PGP-9____   373 GQTPSKISGRISVNKVEFTYPTRADVKILK
                      G  PS ++G+++++ ++FTYPTR DV ILK
                      GVRPSTLTGKLTISNLKFTYPTRPDVPILK
H_S09_0006925__ 22160 ggactacagacaatatatatcaccggcaca
                      gtgccctcgatctcatatcaccgcatctta
                      tgagttcctacacacggtctaaaatgttcg


CAEEL-PGP-9____   403                        GVSLDAQPGQTVALVGSSGCGKSTII
                                             GVS +A PG+T+ALVG SGCGKSTI+
                                             GVSFEANPGETIALVGPSGCGKSTIV
H_S09_0006925__ 22250                        ggttggacggaagtggcagtgaaaag
                      GTGGGAT  Intron 11  TAGgtctacacgactcttgcggggagctt
                      <0-----[22250:22306]-0>tcgcagttgagatgttattcaatcac


CAEEL-PGP-9____   429 QLLQRFYNPDAGQ                          ILIDDIPIEDFNI
                      QLL R+YNPD G+                          I IDD+ I+ +NI
                      QLLLRYYNPDEGK                          ITIDDVEIDKINI
H_S09_0006925__ 22385 cttccttacggga                          aaagggagaaaa
                      atttgaaacaagaGTATTTG  Intron 12  CAGtctaatataatat
                      ggggattcttaag<0-----[22424:22550]-0>tgattgactgcta


CAEEL-PGP-9____   455 KYLRQLVGVVSQEPNLFNTSIEQ                        NIR
                      ++LR  VGVVSQEP LFNT+IEQ                        NIR
                      EFLRNYVGVVSQEPMLFNTTIEQ                        NIR
H_S09_0006925__ 22590 gtccatggggtcgcactaaaagc                        aac
                      attgaatgttcaactttacctaaGTTGGTC  Intron 13  AAGatg
                      actgcccacttagggtctagtag<0-----[22659:22709]-0>ccc


CAEEL-PGP-9____   481 YGRSDVSDEDIARALKEANA                          500
                      YGR +V+D +I  AL++ANA
                      YGRENVTDAEITAALRKANA
H_S09_0006925__ 22719 tgcgagagggaaggccagag                        22778
                      aggaatcacatccctgacac
                      caagtcccgatcttataccc
                                    16 n [22782:22797]


CAEEL-PGP-9____   501 ADFIKTFPEGLNTLVGDRGVQMSGGQKQRIAIARALVRNPKILLLDEATS
H_S09_0006925__       ----------------------------------------------------


CAEEL-PGP-9____   551 ALDAESESIVQSALENASRGRTTIVIAHRLSTVRNADKIIVMKAGQVMEV
H_S09_0006925__       ----------------------------------------------------


CAEEL-PGP-9____   601 GTHETLIEQKGLYHELVHAQVFADVDDKPKKKEAERRMSRQTSQRKGSVN
H_S09_0006925__       ----------------------------------------------------


CAEEL-PGP-9____   651 FKTQESQVDEKPGAPPAPEA                          670
H_S09_0006925__       --------------------


CAEEL-PGP-9____   671 AEKEIKRLKK              E
                      A K+++RLKK              E
                      AVKDLERLKK              E
H_S09_0006925__ 23009 ggagcgccaag           [gaa]
                      ctaatagtaa GTCGCCT  Intron 15    CAGa
                      cagtggagaa <1-----[23040:23094]  -1>a


CAEEL-PGP-9____   682 LEEEGAVKANLFKILRYARPEWIYIFFAIIAALIQGAVMPAFSL
                      LEEEGAV+ANL+KI RYARPEW +I++A++++++QG V+PAFSL
                      LEEEGAVRANLIKIFRYARPEWPFIIIAVLSSIVQGCVFPAFSL
H_S09_0006925__ 23097 cgggggagacaaatatgccgtctaaaggcttagcgtgtcgttt
                      taaagctgcattattgacgcagcttttcttccttaggttcctct
                      caaattcagcctactgttacagttcttcaccacaaacttatcca


CAEEL-PGP-9____   726 FFSQIINVFSN                                   736
                      FF+QII+V+
                      FFTQIIDVLYK
H_S09_0006925__ 23229 ttacaaggcta                                  23261
                      ttcattataa
                      ccaatctagcg


CAEEL-PGP-9____   737 --------PDR             -                    739
                             P
                      FLHSQVLPPSI             C
H_S09_0006925__ 23262 tcctcgccctat          [tgt]                  23327
                      ttacattccct GTTCCAT  Intron 16  CAGg
                      cctcatttact <1-----[23296:23325]-1>t


CAEEL-PGP-9____   740 -DQMKKDGHFWALMFLVLAAVQGTSMLFQ
                        +K DGHFWALMFL+L  +Q  SML+Q
                      DPSLKSDGHFWALMFLLLGVIQAISMLLQ
H_S09_0006925__ 23328 gctcaaggcttgcatcctggacgataccc
                      acctagagatgctttttgttactctttaGTAGCTC  Intron 17
                      tagcgtcttgacgttcgtttatctgtag<0-----[23415:23472]


CAEEL-PGP-9____   768    CSLFGVAAERLTMRIRSKVYRNVLRQDATYFDMPKHSPGRITTRLA
                         C  FG++AERLTMR+RSK++ NV+R DATYFDMP+HSPG+ITTRLA
                         CFFFGMSAERLTMRLRSKIFNNVMRMDATYFDMPRHSPGKITTRLA
                         ttttgatggccaaccctaataagaaaggattgaccctcgaaaaccg
H_S09_0006925__ 23470 CAGgtttgtccagtctgtgcattaattgtaccatatcgaccgatccgtc
                      -0>tccctgacagcaggttagccctggagtcccccgtttatcacccagc


CAEEL-PGP-9____   814 TDAPNIKS                          AIDYRLGSIFNAIASVGG
                      TDAPN+KS                          A+DYR G++F+++ SVG
                      TDAPNVKS                          ALDYRFGAVFSSLVSVGC
H_S09_0006925__ 23611 aggcagat                          gcgtctgggtatcgtggt
                      caccatacGTATGTT  Intron 18  CAGctaagtgcttgcttctgg
                      gtactaat<0-----[23635:24083]-0>ccccttttccttattcgtt


CAEEL-PGP-9____   840 GLGIAFYYGWQMAFL                          VMAIFPFMAVG
                      G+GIAFY GWQMA+L                           ++IFP+ A G
                      GIGIAFYSGWQMALL                          AISIFPLGAAG
H_S09_0006925__ 24138 gagagtttgtcagct                          gatatccgggg
                      gtgtctacggatcttGTGAGTA  Intron 19  CAGctcttctgccg
                      ctatacttagagatg<0-----[24183:24238]-0>acccttgtatg


CAEEL-PGP-9____   866 QALMMKYHGGSATSDAKEMENAGK                        TA
                      +A++M++  G AT+DAKEMEN+GK                         A
                      RAIQMRFMSGRATADAKEMENSGK                        IA
H_S09_0006925__ 24272 cgacaatatgcgagggagagaaga                        ag
                      gctatgttcggcccacaataaggaGTAGATC  Intron 20  CAGtc
                      agtgggtgtgttactggggattgg<0-----[24344:24401]-0>tt


CAEEL-PGP-9____   892 MEAIENIRTVQALTLQTKLYNIFCSHLDAPHGGNISKAIIR
                      MEAIE+IRTVQALTL+ +L+  FC HLD PH  N  KA+++
                      MEAIEHIRTVQALTLERRLHAEFCHHLDGPHETNRRKALLQ
H_S09_0006925__ 24408 aggagcacagcgtatgccccggttcccggccgaaacagccc
                      tactaatgctactctaggtacatgaatagcaacaggactta
                      gaccgttagggagggatgttcactttgtaacaccagacttg


CAEEL-PGP-9____   933                          GLTYGFANSIQFFTYAAAFRFGLFLI
                                               G++YGFA SI +F YA+ FRFG +LI
                                               GVSYGFACSIPYFLYASCFRFGAWLI
H_S09_0006925__ 24531                          ggttgtgtaatttttgtttctggtca
                      GTCAGGC  Intron 21  CAGgtcagtcggttattaccgtgtggtt
                      <0-----[24531:24619]-0>ccactctcctctcactaccataaggc


CAEEL-PGP-9____   959 FDKNVLMEPENVL              R          VLFAISFSFG
                            + P NVL              R          V +AISF+ G
                      V--HGYLGPMNVL              R          VELAISFTAG
H_S09_0006925__ 24698 g  cgtcgcaagca           [agg]        ggcgattagg
                      t  agatgctattgGTAAGAA  Intron 22  CAG tatctctccg
                      c  ccccgggctc <2-----[24733:24850]-2>gtgcattctta


CAEEL-PGP-9____   983 TIGFAASYFPEYIKATFAAGLIFNMLEEEPRIDGMTSSGTYPQLSGEVK
                      ++GFA+SYFPEYIKATFAAGLIF+MLEE+PRIDGM+++G  P+++G VK
                      SLGFASSYFPEYIKATFAAGLIFHMLEEKPRIDGMSNNGKKPKITGAVK
H_S09_0006925__ 24882 acgtgatttcgtaagatgggcatcacggaccaggaaaagaacaaaaggga
                      gtgtcgcatcaatacctccgtttattaaacgtagtgaagaacatcgcta
                      ctttccgtccgccatattctcttcggaaggcccgccccaagacccca


CAEEL-PGP-9____  1032 LNKVFFRVPERPAVPILQGLNVH                        VKP
                      L++V+F YPERP VPILQGLN+                        V+P
                      LDQVYFNYPERPDVPILQGLNIK                       VEP
```

```
H_S09_0006925__25029 cgcgttatcgccggcaccgcaaa                    ggc
                     taatataacagcatcttagtataGTGAGTA  Intron 23  CAGtac
                     gcatccttcgaatgcatgactcg<0-----[25098:25349]-0>gac


CAEEL-PGP-9____ 1058 GQTLALVGPSGCGKSTVISLLERLYDPLEGAV
                     G+TLALVGPSGCGKSTVISLLER YD L+G+V
                     GETLALVGPSGCGKSTVISLLERFYDALDGSV
H_S09_0006925__25359 ggacgcggcagtgatagatccgattggtggtg
                     gactcttgcggggaccttcttagtaactagct
                     cgcgtgtaatttagaaaagatagtcttgcctt


CAEEL-PGP-9____ 1090                      TVDNNDLRQMNPKHLRKHIALVSQEP
                                           VD NDLR+MNP HLR HIALVSQEP
                                          EVDGNDLREMNPTHLRAHIALVSQEP
H_S09_0006925__25455                      ggggagtcgaacacccgcagtgtcgc
                     GTAAGTC  Intron 24  CAGatagaatgataccatgcatcttcaac
                     <0-----[25455:25858]-0>acttccgtagctacgtcttagaagga


CAEEL-PGP-9____ 1116 ILFDTSIRENIVYGLQPGEYTHEQITECSKANIHKFIDELPD
                     ILFD SIR+NI+YGL PG  +  ++     +ANIHKF+ ELP+
                     ILFDRSIRDNILYGLPPGSVSEAEVHEVAQRANIHKFVMELPE
H_S09_0006925__25937 actgatacgaactgtccgtgaggggcgggccgaacatgagtcg
                     tttagctgaattagtccgctgacataatcagcataatttatca
                     ccccaccactcctcgaatcctatagtaccattctcatagagcg


CAEEL-PGP-9____ 1159                      GYETRVGEKGTQLSGGQKQRIAIARA
                                          GY TR GEKG QLSGGQKQRIAIARA
                                          GYNTRAGEKGVQLSGGQKQRIAIARA
H_S09_0006925__26066                      gtaacgggaggcttggcaccagagcg
                     GTGTGTT  Intron 25  TAGgaacgcgaagtatcggaaagtctcgc
                     <0-----[26066:26118]-0>cccatgaaaccggttggaaacccctc


CAEEL-PGP-9____ 1185 LIRNPKILLLDEATSALDTESEK                        QVQ
                     LIRNPKILLLDEATSALDTESEK                         VQ
                     LIRNPKILLLDEATSALDTESEK                        VVQ
H_S09_0006925__26197 caaacaatccgggaagtgagaga                         ggc
                     ttgacattttaaccgctacagaaGTACTAT  Intron 26  CAGtta
                     ccatgatggtcgtctcgtcacgg<0-----[26266:26342]-0>tga


CAEEL-PGP-9____ 1211 VALDAAAKDRTCIVVAHRLSTIVNAGCIMVVKNGQVVEQ
                      ALD A++ RTCIVVAHRLST+VNA CIMVV+ G++VE+
                     EALDKASEGRTCIVVAHRLSTVVNANCIMVVQGGKIVEK
H_S09_0006925__26352 ggcgagtggcatagggccctaggagataaggcggaagga
                     actaaccaggcgtttcagtccttacagttttaggattaa
                     ggccgaaaccctctgatgggtcctctccgacaaaattaa


CAEEL-PGP-9____ 1250              G          THNELIAKRGAYFALTQKQSSNQ
                                  G          THNEL+  +G Y+ LTQKQ++
                                  G          THNELMQAKGVYWELTQKQTT--
H_S09_0006925__26469 g          [gga]         acagcacgaggttgcacacaa
                      GTACAGC  Intron 27  CAGgcaaattacagtagatcaaacc
                      <1-----[26470:26530]-1>aaccaagaaatgtggcaagggg


CAEEL-PGP-9____ 1274 SGGAFDTSEALDDDDDDHVKF                          1294
                                          K
                     ------------------AKE
H_S09_0006925__26596                    gag                        26604
                                        caa
                                        aaa


CAEEL-PGP-9____      *
                     *
                     *
H_S09_0006925__26605 t                                              26607
                     a
                     a


H_S09_0006925__26608 [26608:27070]                                  27070
//

Gene 1
EXONS 16713 26604
  Exon 16713 16783 phase 0
  Exon 18125 18161 phase 2
  Exon 18351 18465 phase 0
  Exon 18518 18571 phase 1
  Exon 19514 19602 phase 1
  Exon 20637 20770 phase 0
  Exon 20830 21001 phase 2
  Exon 21110 21234 phase 0
  Exon 21673 21841
       218482 21848 phase 2
  Exon 21911 22024 phase 0
  Exon 22130 22249 phase 0
  Exon 22307 22423 phase 0
  Exon 22551 22658 phase 0
  Exon 22710 22778 phase 0

  ---

  Exon 23009 23039
  Exon 23095 23261
       23262 23295 phase 1
  Exon 23326 23327
       23328 23414 phase 1
  Exon 23473 23634 phase 0
  Exon 24084 24182 phase 0
  Exon 24239 24343 phase 0
  Exon 24402 24530 phase 0
  Exon 24620 24732 phase 0
  Exon 24851 25097 phase 2
  Exon 25350 25454 phase 0
  Exon 25859 26065 phase 0
  Exon 26119 26265 phase 0
  Exon 26343 26469 phase 0
  Exon 26531 26604 phase 1
//

Making a K in phase 2 intron
Making a T in phase 1 intron
Making a T in phase 1 intron
Making a D in phase 2 intron
Making a R in phase 2 intron
Making a E in phase 1 intron
Making a C in phase 1 intron
Making a R in phase 2 intron
Making a G in phase 1 intron


>H_S09_0006925___HAECO_S09_Supercontig_0006925
_16713-21841bp_AA
_21842-21848bp_Frameshift
_21849-22778bp_AA
---
_23009-23261bp_AA
_23262-23327bp_AA
_23328-26604bp_AA
_CAEEL-PGP-9_1-323AA_324AA_Frameshift_325-500AA---671-736AA_737-
739AA_740-1294AA_End

MGFLKKKGKAADSKESENDNEEEKKDDAPKASIPQLFRYTTTFDKLLLLIGSTVAMGTGM
GLPMMAIIMVRHEFTTITYGLDASSTTAPYRLTFQQFEHDVTQNCLKYVYLGCGIFAAAT
IQATCFLTVGENLVNQLRRQFFKSILRQDIPWFDKNGSGTLATKLFDNLERVKEGTGDKV
GLMIQYVAQFFGGFIVAFTYDWKLTLIMMSLAPFMIMCGAFIARLMATAATREAKKYAVA
GGIAEEVLTSMRTVMAFNGQPYECERYKQALEDGKSTGIKKSLYVGIGFAITFLILFSSY
CLAFWVGTDFVYNNRMQGGTVM
```

```
TA


V
FFAVMMGSMALGQAGPQFAVLGTAMGAAGSLYQI
IDREPEIDSYSQEGVRPSTLTGKLTISNLKFTYPTRPDVPILKGVSFEANPGETIALVGP
SGCGKSTIVQLLLRYYNPDEGKITIDDVEIDKINIEFLRNYVGVVSQEPMLFNTTIEQNI
RYGRENVTDAEITAALRKANA

AVKDLERLKKE
LEEEGAVRANLIKIFRYAR
PEWPFIIIAVLSSIVQGCVFPAFSLFFTQIIDVLYK

FLHSQVLPPSIC

DPSLKSDGHFWA
LMFLLLGVIQAISMLLQCFFFGMSAERLTMRLRSKIFNNVMRMDATYFDMPRHSPGKITT
RLATDAPNVKSALDYRFGAVFSSLVSVGCGIGIAFYSGWQMALLAISIFPLGAAGRAIQM
RFMSGRATADAKEMENSGKIAMEAIEHIRTVQALTLERRLHAEFCHHLDGPHETNRRKAL
LQGVSYGFACSIFYFLYASCFRFGAWLIVHGYLGPMNVLRVELAISFTAGSLGFASSYFP
EYIKATFAAGLIFHMLEEKPRIDGMSNNGKKPKITGAVKLDQVYFNYPERPDVPILQGLN
IKVEPGETLALVGPSGCGKSTVISLLERFYDALDGSVEVDGNDLREMNPTHLRAHIALVS
QEPILFDRSIRDNILYGLPPGSVSEAEVHEVAQRANIHKFVMELPEGYNTRAGEKGVQLS
GGQKQRIAIARALIRNPKILLLDEATSALDTESEKVVQEALDKASEGRTCIVVAHRLSTV
VNANCIMVVQGGKIVEKGTHNELMQAKGVYWELTQKQTTAKE
//

>H_S09_0006925___HAECO_S09_Supercontig_0006925
_16713-21841bp_AA
_21842-21848bp_AA
_21849-22778bp_AA
---
_23009-23261bp_AA
_23262-23327bp_AA
_23328-26604bp_AA

atgggtttcttaaagaagaaagggaaagcagccgatagcaaagagtcagagaatgacaaac
gaagaagaaaaagaaagatgatgcaccgaaagccagtatcccacaactgttccgttacaca
accacattcgacaagttgttgctcttgatcggttcaactgtcgctatgggtactggtatg
ggactcccaatgatggctattatcatggtaaggcatgaattcactaccataaacctacgga
ttggacgcaagtagcaccaccgctccatacaggcttacttttcaacaatttgaacacgac
gtgacacaaaactgtttgaaatacgtctaccttggttgtggaatattcgcagcagcaacg
attcaggcgacatgtttcctgactgtcggagaaaacctggtgaatcaattgaggcgacag
tttttcaagtcaattcttcgccaagacattccatggtttgataagaatggttcaggaaca
ttggccacaaagctgttcgacaatctcgaacgagtcaaggaaggaaccggcgacaaagtc
ggtctgatgatccagtatgtggcacagttcttcggtggcttcatcgtggcattcacctat
gactggaaactcacgctgattatgatgtccctggctccattcatgatcatgtgcggagca
tttattgctaggttgatggccactgcagcgactcgtgaggccaaaaagtatgcagtggct
ggaggaatagccgaagaggtgcttacttcgatgaggactgttatggctttcaacggacag
ccttatgaagtgcgaaaggtacaaagcaggcactggaaagatggttaaatctacaggaataaag
aaatccttgtatgttggtattggcttcgcaattacatttctgatcttgttctcttcctat
tgcttggcatttgggtcggcacggatttcgtctacaataaccggatgcaaggtggaact
gtgatg

acggca

a

gtc
ttttttgccgttatgatgggctctatggcgcttggacaagctggacctcagtttgctgta
cttggtacagctatgggtgcagccggatctctctatcagatcatcgatagggaaccagag
atcgactcctactcacaggaaggtgtgagaccgtctactctcaccggtaaactcacaatc
tcaaacttgaagtttacctatccaacacgaccagatgtgcctattctcaagggtgtctcg
ttcgaagcgaatcctggggaaacgatagctttggttggtccaagtggttgcggaaaaagt
accatagtccagttgttgctgcgatattataaccctgatgaaggaaagattacgatagat
gatgtggaaatcgataagatcaatatagaattccttcggaactacgtcggagtcgtttct
caagagccgatgcttttcaatacaacgattgaacagaacatccgctacggacgagagaat
gtcaccgacgcggaaattaccgctgctctacgtaaagccaacgcc

gccgtaaaggatctggagcgactgaaaaaagaactcgaa
gaagaaggtgctgtcagagcgaacctcattaaaatctttaggtatgctcgacccgaatgg
ccttttatcattattgccgtactctcctcaatcgtacaaggatgcgtttttccagctttc
tcctattcttcacacaaattatcgatgtactgtacaag

ttcctccattcccaagttctt
cctccatccatttgt

gatccatcgctcaagagcgatggccatttttgggcactcatgttt
cttctcttgggtgttattcaagctatctctatgcttctacagtgtttcttcttcggtatg
tcagccgaacggctcacaatgcggcttcgttcaaagatcttgatgagaatg
gatgccacctacttcgacatgcctcgtcattcacctggcaaaatcaccacccgactggcc
acggatgcacccaatgtaaaatctgccctcgactaccgttttggtgccgtctttagttca
cttgtttccgtgggttgtggcattggaattgcattctattctggatggcaaatggcactt
ttggcaatctccatcttctcctcgggtgcagctgggcgagcgattcagatgaggtttatg
tctggggctgctacagccgatgcgaaggagatggaaaatagtgggagattgctatggaa
gccatcggagcatattcgaacggtgcaggcattgacgttggaacgtcggcttccatgccgaa
ttctgtcatcatctggatggacccacacgaaaccaacagacggaaagcccttcttcagggc
gtctcatacggtttcgcttgcagcattttctatttcttatacgcttcatgcttccgattt
ggagcatggctgatcgtccacggctacctcgggccgatgaacgttctcagggttgagctc
gcaatttctttcactgctggaagccttggttttgccagctcgtatttccccgagtacatc
aaagctacatttgctgccggtctcatttttcacatgctggaagaaaagccgcgcatcgac
ggcatgagcaacaacggcaaaaaaccgaaaatcaccggcgccgtcaaactggaccaagtt
tacttcaattatcccgagcgaccagatgtgcccatacttcagggactcaatatcaaggtg
gaacccggcgagaccctggctctggttggaccaagtggttgtggaaagtcaacagtaata
tcgctacttgaaaggttttacgatgctttggacggctctgttgaagtcgatggtaacgac
ttgcgtgaaatgaaccctacacacctgcgtgcccatattgcattggtatcacaggagcca
atcctcttcgacagatccatccgagacaatatcctctatggcttgccaccaggttccgtc
agtgaagctgaagtgcatgaagtcgccccaacgtgctaacattcacaaatttgtaatggaa
ttgcccgagggctacaacacacgtgcgggagaaaaaggcgtccagttgtctggtgggcag
aaacaacgaatcgccatcgcccgtgccctcatcagaaatccgaaaattttgctgcttgac
gaggctaccagtgccttggataccgaaagcgagaaggttgtgcaagaggcgctcgacaag
gcatcagaaggccgcacctgtatcgttgtggcacatcggctgtcgactgtcgtcaatgcc
aattgcatcatggtagtccaaggaggaaaaattgttgaaaaaggaacacacaacgaacta
atgcaagcaaaaggtgtgtattgggagctcacacaaaagcagacgacggcaaaagaa
//

genewise

H_S09_0006925_-27070          Intron     CAG
                     <0-----[     : 2005]-0>
                       16 n [22797:22782]
                       13 n [19842:19830]
                      328 n [17365:17038]
                       10 n [15190:15181]
                       16 n [11423:11408]
                       10 n [4070 : 4061]
                       10 n [2045 : 2036]

CAEEL-PGP-9____  557 ESIVQSALENASRGRTTIVIAHRLSTVRNADKIIVMKAGQ
                      SI ++ +  AS+GRTTIV+AHRLST+RNAD+II +K G+
                     ISIRRTEVIQASKGRTTIVVAHRLSTIRNADRIIAIKDGE
H_S09_0006925__-2004 ataccaggacgtagaaaagggccttcaacaggaaagaaggg
                     tctggcattaccaggcctttcagtcctgacagttctaaga
                     acagatagtgtcatatcccttaggttattcactccataa


CAEEL-PGP-9____  597                                  VMEVGTHETLI
                                                      V+EVGTH+ L+
                                                      VAEVGTHDELM
```

```
H_S09_0006925__-1884                                    gggggacggca
                    GTACTGCAGTTCGAG  Intron 1   TTAAGtcatgcaaatt
                    <0-------------[1884 : 1824]---0>ggaccttcgtg


CAEEL-PGP-9____   608 EQKGLYHELVHAQVFADVDD              K          PKK
                       +KGLY ELV+AQVF DVD
                    ARKGLYRELVNAQVFVDVDG              A          ---
H_S09_0006925__-1790 gcagctcgtgagcgtgggggg           [gca]
                    cgagtagattacatttatag GTACGCT  Intron 2   GAGc
                    ctaagctagcttggcctcca <1-----[1729 : 1096]-1>a


CAEEL-PGP-9____   632 KEAERRMSRQTSQRKGSVNFKTQESQVDEKPGAPPAPEAAEKEIKRLKK
                                Q  GS  + T + +  EK  A  + +A  K+++RLKK
                    --------YDVFQGLGS-QLSTSDDK-REKNAANKSVKAV-KDLERLKK
H_S09_0006925__-1093         tggtcgcgt cctaagga agaaggaatgagg agcgccaa
                             aattagtgc atccgaaa gaaaccaactact aataqtaa
                             tcgtgccac aatgtcca aaactgcgggaca gcgaagga


CAEEL-PGP-9____   681              E          LEEEGAVKANLFKILRYARPEWI
                                   E          LEEEGAV+ANL+KI RYARPEW
                                   E          LEEEGAVRANLIKIFRYARPEWP
H_S09_0006925__ -979 g          [gaa]        cgggggagacaaatatgccgtc
                     GTCACCT  Intron 3   CAGataaagctgcattattgacgcagc
                     <1-----[978  :  924]-1>acaaattcagcctactgttacagt


CAEEL-PGP-9____   705 YIFFAIIAALIQGAVMPAFSLFFSQIINVFSNPDRDQM
                    +I +A++++++QG V+PAFSLFF+QII+V+
                    FITIAVLSSVVQGCVFPAFSLFFTQIIDVL-------Y
H_S09_0006925__ -852 taaaggcttggcgtgtcgtttttacaaggc       t
                    ttctcttccttaggttcctctttcattatt       a
                    tcttcaccacaaacttatccaccaatccag       c


CAEEL-PGP-9____   743              K          KDGHFWALMFLVLAAVQGTSMLF
                                              DGHFWALMFLVL  +Q  SML+
                                   T          SDGHFWALMFLVLGIIQAISMLL
H_S09_0006925__ -759 a           [acg]        aggcttgcatcgcgaacgatacc
                     cGTTCCTC  Intron 4   CAG gagatgctttttttgttactcttt
                      <2-----[757  :  677]-2>gctcttgacgtcagtttatctgta


CAEEL-PGP-9____   767 Q                     CSLFGVAAERLTMRIRSKVYRNVLR
                      Q                     C  FG++AERLTMR+RSK++ NV+R
                      Q                     CFFFGMSAERLTMRLRSKIFNNVMR
H_S09_0006925__ -606 c                     ttttgatggccaaccctaataagaa
                     aGTAGCTG  Intron 5   CAGgtttgtccagtctgtgcattaattg
                     g<0-----[603  :  551]-0>tccctgacagcgggttagccctgga


CAEEL-PGP-9____   793 QDATYFDMPKHSPGRITTRLATDAPNIKS                 821
                       DATYFDMP+HSPG+ITTRLATDAPN+KS
                      MDATYFDMPRHSPGKITTRLATDAPNVKS
H_S09_0006925__ -475 aggattgaccctcgaaaaccgaggcagat                 389
```

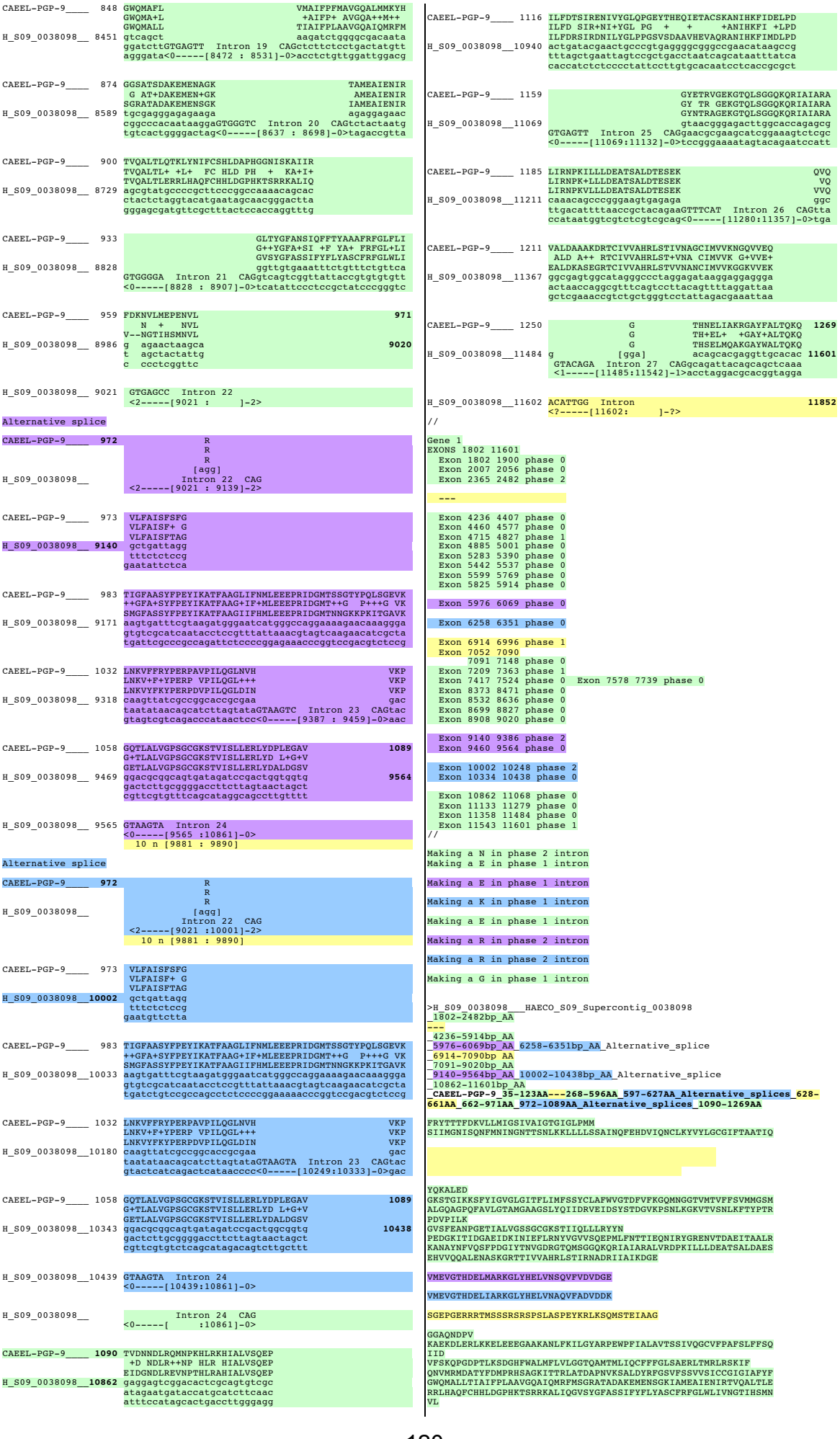
```
                     taccatatcgaccgatccgtccaccatac
                     gtcccctgtttatcaacgagcgtactgat


H_S09_0006925__ -388 GTACGTT  Intron                                  1
                     <0-----[388  :     ]-0>
//

Gene 1
EXONS 2004 389
  Exon 2004 1885 phase 0
  Exon 1823 1730 phase 2
  Exon 1095 979 phase 1
  Exon 923 758 phase 1
  Exon 676 604 phase 2
  Exon 550 389 phase 0
//

Making a A in phase 1 intron
Making a E in phase 1 intron
Making a T in phase 2 intron


>H_S09_0006925___HAECO_S09_Supercontig_0006925
_2004-389bp_AA
_CAEEL-PGP-9_557-821AA

E
SIRRTEVIQASKGRTTIVVAHRLSTIRNADRIIAIKDGE
VA
EVGT
HDELMARKGLYRELVNAQVFVDVDGAYDVFQGLGSQLSTSDDKREKNAANKSVKAVKDLE
RLKKELEEEGAVRANLIKIFRYARPEWPFITIAVLSSVVQGCVFPAFSLFFTQIIDVLYT
SDGHFWALMFLVLGIIQAISMLLQCFFFGMSAERLTMRLRSKIFNNVMRMDATYFDMPRH
SPGKITTRLATDAPNVKS
//

>H_S09_0006925___HAECO_S09_Supercontig_0006925
_2004-389bp

ata
tccatacggcgaactgaagtgattcaggcttccaaaggtagaactaccatcgtcgttgct
catcgattgtcgactattcgaaatgctgacagaatcattgccatcaaagatggagaa
gtggcg
gaagtcggcact
catgacgagcttatggcccgtaaaggactgtaccgtgaattggtcaatgctcaggtgttc
gtcgatgtcgacggagcatatgacgtgtttcagggcctcggatcccaactatctacgagt
gacgacaaaagagaaaaaaacgctgcgaacaagtcggtgaaagccgtaaaggacctggaa
cgactgaagaaagaactcgaagaagaaggtgctgtcagagcgaacctcattaaaatcttt
aggtatgctcgacccgaatggccttttatcactattgccgtactctcctcagtcgtacaa
ggatgcgtttttccagctttctccttattcttcacacaaattatcgacgtactgtacacg
agcgatggccatttttgggcactcatgtttctcgtactgggtattattcaagctatctct
atgcttctacagtgtttcttcttcggtatgtcagccgaacggctcacgatgcggcttcgt
tcaaagatcttcaacaatgtgatgagaatggatgccacctacttcgatatgcctcgtcat
tcacctggcaaaataaccacgcgactggccacggatgcacccaatgtgaaatct
//
```

A.24 *HAECO*_S09_SUPERCONTIG_0038098 ≈ *CAEEL*-PGP-9

```
genewise
Query protein:        CAEEL-PGP-9____CAEEL-PGP-9_C47A10_1_1-1294AA
Comp Matrix:          blosum62.bla
Gap open:             12
Gap extension:        2
Start/End             global
Target Sequence       H_S09_0038098___HAECO_S09_Supercontig_0038098
Strand:               both
Start/End (protein)   global
Gene Paras:           worm.gf
Codon Table:          codon.table
Subs error:           1e-05
Indel error:          1e-05
Model splice?         model
Model codon bias?     flat
Model intron bias?    tied
Null model            syn
Algorithm             623
```

```
H_S09_0038098__     1          Intron      CAG
                    <0-----[      : 1801]-0>


CAEEL-PGP-9____    35 YRYTSTVDRLMLAVGI
                      +RYT+T D+++L +G
                      FRYTTTFDKVLLMIGS
H_S09_0038098__  1802 tctaaatgagccaagt
                      tgaccctaatttttgc
                      cccgtattgggcgtca


CAEEL-PGP-9____    51 IVSCATGVGLPLMSIIM                          GNVSQNFVT
                      IV+  TG+GLP+MSIIM                          GN+SQNF+
                      IVAIGTGIGLPMMSIIM                          GNISQNFMN
H_S09_0038098__  1850 aggagagagccaataaa                          gaatcataa
                      ttctgcgtgtcttctttGTAAGTC  Intron 1   CAGgatcaatta
                      tctttttaataggtttg<0-----[1901 : 2006]-0>tttaaccgt


CAEEL-PGP-9____    77 LGTIFLD              P          NSTASEKAAARAEFSH
                      +                               ++A  +F H
                      INGNTTS              N          LKKLLLLSSAINQFEH
H_S09_0038098__  2034 aagaaaaa           [aat]        taactctttgaactgc
                      tagaccgaGTGAGCT  Intron 2   CAG taattttccctaataa
                      ctacttt <2-----[2057 : 2364]-2>taggtatatagctatat


CAEEL-PGP-9____   101 EVIQNCLKYVYLGCGIFAAGFLQ                       123
                      +VIQNCLKYVYLGCGIF A  +Q
                      DVIQNCLKYVYLGCGIFTAATIQ
H_S09_0038098__  2414 ggacattatgtcgtgataggaac                     2482
                      attaagtaatatgggttccccta
                      tgcgttgacctcctgacgtcaag


H_S09_0038098__  2483 GTATGGT  Intron 3
                      <0-----[2483 :     ]-0>
                       990 n [2886 : 3875]


CAEEL-PGP-9____   124 ASCFMVICEKLSNRFRRQFFHSVMRQEIAWYDKNTSGTLSNKLFDNLERV
H_S09_0038098__       ------------------------------------------------------


CAEEL-PGP-9____   174 REGTGDKVGLAFQMMAQFIGGFAVAFTYDWLLTLIMMSLSPFMMICGLFL
H_S09_0038098__       ------------------------------------------------------


CAEEL-PGP-9____   224 AKLLATAATKEAKQYAVAGGIAEEVLTSIRTVIAFNGQEYECKR   267
H_S09_0038098__       ---------------------------------------------


H_S09_0038098__          Intron 4   TTCAG
                      <2-----[     : 4235]---2>


CAEEL-PGP-9____   268  YEDALEHGKK
                       Y+ ALE GK
                       YQKALEDGKS
H_S09_0038098__  4236 gtcagtgggat
                       aaactaagac
                       caatgactag


CAEEL-PGP-9____   278 TGIKKSFLIGAGLASFFVIIYASYCLAFWVGTNFVYSGRLESGTVLT
                      TGIKKSF IG GL   F+I+++SYCLAFWVGT+FV+ G++  GTV+T
                      TGIKKSFYIGVGLGITFLIMFSSYCLAFWVGTDFVFKGQMNGGTVMT
H_S09_0038098__  4267 agaaatttagggcgaatcaatttctcgttggagtgtagcaaggagaa
                      cgtaactatgtgtgtcttttcagtctgtgcatttagataggcttc
                      gacgattttcttgaagcgcgccgtcgatgccgctccatagtaatgga


CAEEL-PGP-9____   325                          VFFSVMMGSMALGQAGQQFATIGTAL
                                               VFFSVMMGSMALGQAG QFA +GTA+
                                               VFFSVMMGSMALGQAGPQFAVLGTAM
H_S09_0038098__  4408                          gtttgaagtagcgcggcctggcgaga
                      GTACGTT  Intron 5   AAGtttctttgctctgacgcatcttgcct
                      <0-----[4408 : 4459]-0>tcccaggtcggcaatatgttcctgtg


CAEEL-PGP-9____   351 GAAASLYEVIDRI                P          EIDAYSTEGQ
                      GAA SLY++IDR+                            ID+YST+G
                      GAAGSLYQIIDRV                E          -IDSYSTDGV
H_S09_0038098__  4538 ggggtctcaagcgg                [gaa]        agtttaggg
                      gccgctaattagt GTTCAAT  Intron 6   CAGa tacaccagt
                      tacatctgtttga <1-----[4578 : 4714]-1>a ctccagtcg


CAEEL-PGP-9____   375 TPSKISGRISVNKVEFTYPTRADVKILK
                       PS + G+++V+ ++FTYPTR DV ILK
                      KPSNLKGKVTVSNLKFTYPTRPDVPILK
H_S09_0038098__  4744 actacagagagtacatatcaccggcaca
                      accatagatctcatatcaccgcatcttaGTTTGCC  Intron 7
                      aagtcagaaacatgacttagaatttttg<0-----[4828 : 4884]


CAEEL-PGP-9____   403   GVSLDAQPGQTVALVGSSGCGKSTIIQLLQRFYNPDAGQ
                        GVS +A PG+T+ALVGSSGCGKSTIIQLL R+YNP+ G+
                        GVSFEANPGETIALVGSSGCGKSTIIQLLLRYYNPEDGK
H_S09_0038098__  4885   ggttggacggaagtggtagtgaaaaacttccttacggga
                      CAGgtctacacgactcttgcggggagcttatttgaaacaaga
                      -0>acacaatttagaaggtctttaatcacggagatcttatag


CAEEL-PGP-9____   442                          ILIDDIPIEDFNIKYLRQLVGVVSQE
                                               I ID   I+ +NI++LR  VGVVSQE
                                               ITIDGAEIDKINIEFLRNYVGVVSQE
H_S09_0038098__  5002                          aaaggggagaaaagtccatggggtcg
                      GTATGCC  Intron 8   TAGtctagcataatatattgaatgttcaa
                      <0-----[5002 : 5282]-0>cgattgaccgtctgctacccaccaag


CAEEL-PGP-9____   468 PNLFNTSIEQ                          NIRYGRSDVSDEDIAR
                      P LFNT+IEQ                           NIRYGR +V+D +I
                      PMLFNTTIEQ                           NIRYGRENVTDAEITA
H_S09_0038098__  5361 cactaaaagc                           aactgcgagagggaag
                      ctttacctaaGTTGGTC  Intron 9   CAGatgaggaatcacatcc
                      ggtctactag<0-----[5391 : 5441]-0>ccgcaagtcctgacct


CAEEL-PGP-9____   494 ALKEANAADFIKTFPE                        GLNTLVGDRG
                      AL++ANA +F+++FP+                        G+ T VGDRG
                      ALRKANAYNFVQSFPD                        GIYTNVGDRG
H_S09_0038098__  5490 gccagagtatgcttcg                        gataagggcg
```

```
                      ctgacacaattactcaGTAAGTG  Intron 10  CAGgtacatgagg
                      tataactcctgagttt<0-----[5538 : 5598]-0>tccgtctctt


CAEEL-PGP-9____   520 VQMSGGQKQRIAIARALVRNPKILLLDEATSALDAESESIVQSALEN
                       QMSGGQKQRIAIARALVR+PKILLLDEATSALDAESE +VQ ALEN
                      TQMSGGQKQRIAIARALVRDPKILLLDEATSALDAESEHVVQQALEN
H_S09_0038098__  5629 acatggcaccagagcgtgagcaactcgggaagcgggagcggccgcga
                      catcggaaagtctcgcttgacattttaaccgctacagaattaactaa
                      cggctcgggtcctttggcatgatcggtgctctcccatatccggtgac


CAEEL-PGP-9____   567                          ASRGRTTIVIAHRLSTVRNADKIIVM
                                               AS+GRTTIV+AHRLST+RNAD+II +
                                               ASKGRTTIVVAHRLSTIRNADRIIAI
H_S09_0038098__  5770                          gtagaaaagggccttaacaggaaaga
                      GTAAGAA  Intron 11  CAGccaggcctttcagtcctgacagttct
                      <0-----[5770 : 5824]-0>ttaaatcccactaggttattcactct


CAEEL-PGP-9____   593 KAGQ                                          596
                      K G+
                      KDGE
H_S09_0038098__  5903 aggg                                        5914
                      aaga
                      ataa


H_S09_0038098__  5915 GTACTGC  Intron 12
                      <0-----[5915 :     ]-0>


Alternative splice

H_S09_0038098__          Intron 12  TTTTTGAG
                      <0-----[5915 : 5975]------0>


CAEEL-PGP-9____   597 VMEVGTHETLIEQKGLYHELVHAQVFADVDD                627
                      VMEVGTH+ L+  +KGLYHELV++QVF DVD
                      VMEVGTHDELMARKGLYHELVNSQVFVDVDG
H_S09_0038098__  5976 gagggacggcagcagctcgtgatcgtgggggg               6069
                      ttatgcaaattcgagtaaattacatttatag
                      ggaccttcgggcgaagctagcttggcctccg


CAEEL-PGP-9____                -

                               E
                             [gaa]
H_S09_0038098__          GTGAGCT  Intron 13
                         <1-----[6070 : 6913]---1>
                           10 n [6171 : 6180]


Alternative splice

H_S09_0038098__          Intron 12  TTCCTTTTCAG
                      <0-----[5915 : 6257]---------0>
                        10 n [6171 : 6180]


CAEEL-PGP-9____   597 VMEVGTHETLIEQKGLYHELVHAQVFADVDD                627
                      VMEVGTH+ LI +KGLYHELV+AQVFADVDD
                      VMEVGTHDELIARKGLYHELVNAQVFADVDD
H_S09_0038098__  6258 gagggacggcagcagctcgcgagcgtgggggа               6351
                      ttatgcaaattcgagtaaattacattcataa
                      tgaccattggcacgacctggctcggcccgcc


CAEEL-PGP-9____                -

                               K
                             [aaa]
H_S09_0038098__          GTAAGTT  Intron 13
                         <1-----[6352 : 6913]---1>


H_S09_0038098__          Intron 13  TTAAG
                         <1-----[     : 6913]---1>


CAEEL-PGP-9____   628                          --KPKKKEAERRMSRQTSQRKGS
                                                        P ++     SR  S    S
                                               SGEPGERRRTMSSSRSRSPSLAS
H_S09_0038098__  6914                          aggcggcccaatttctatcttgt
                                               aggacgagggctcccgcgccctcc
                                               atcggtaagtggatatgatagata


CAEEL-PGP-9____   649 VNFK                          ----TQESQVDEK       661
                        +K                                 +++
                      PEYK                          RLKSQMSTEIAAG
H_S09_0038098__  6985 cgta                          ccatcatagaggg      7090
                      caaaGTATAGG  Intron 14  CAGgtacatccatccg
                      catg<0-----[6997 : 7051]-0>gcatagatatttt


CAEEL-PGP-9____   662 PGAPPAPEAAEKEIKRLKK              E          LEEE
                       GA   P  AEK+++RLKK              E          LEEE
                      GGAQNDPVKAEKDLERLKK              E          LEEE
H_S09_0038098__  7091 gggcagcgaggagtgccaag           [gaa]        cggg
                      ggcaaactacaaatagtaa GTGATTT  Intron 15  CAGataaa
                      attgctagaagacgaagga <1-----[7149 : 7208]-1>agaga


CAEEL-PGP-9____   686 GAVKANLFKILRYARPEWIYIFFAIIAALIQGAVMPAFSLFFSQIIN
                      GA KANLFKIL YARPEW +I +A+ ++++QG V+PAFSLFFSQII+
                      GAAKANLFKILGYARPEWPFIALAVTSSIVQGCVFPAFSLFFSQIID
H_S09_0038098__  7223 gggagactaacgtgccgtctagcggattagcgtgtcgttctttcaag
                      gccacattattgacgcagcttctctcccttaggttcctctttcatta
                      tccagttcactcctatagttcccctctactacttcaatactcaacct


CAEEL-PGP-9____   733                          VFSN-PDRDQMKKDGHFWALMFLV
                                               VFS  P    +K DGHFWALMFLV
                                               VFSKQPGDPTLKSDGHFWALMFLV
H_S09_0038098__  7364                          gttaccggcacaaggcttgcatcg
                      GTACTGTTC  Intron 16  CAGttcaacgacctagagatgcttttt
                      <0-------[7364 : 7416]-0>gtcggattggcgtcctcgatgccg


CAEEL-PGP-9____   756 LAAVQGTSMLFQ                          CSLFGVAAERLTMR
                      L   Q  +ML+Q                          C  FG++AERLTMR
                      LGGTQAMTMLIQ                          CFFFGLSAERLTMR
H_S09_0038098__  7489 cggacgaaacac                          ttttgttggccaac
                      tggcactcttaGTATACC  Intron 17  CAGgtttgtccagtctg
                      gtctgggcgctg<0-----[7525 : 7577]-0>cccctgacagcaga


CAEEL-PGP-9____   782 IRSKVYRNVLRQDATYFDMPKHSPGRITTRLATDAPNIKS
                      +RSK+++NV+R DATYFDMP+HS G+ITTRLATDAPN+KS
                      LRSKIFQNVMRMDATYFDMPRHSAGKITTRLATDAPNVKS
H_S09_0038098__  7620 tctaatcagaaaggattgaccctggaaaactgaggcagat
                      tgcattaattgtaccatatcgaccgatccgtccaccatac
                      aaaatcatggagtcccccgtctataactccggataatgag


CAEEL-PGP-9____   822                          AIDYRLGSIFNAIASVGGGLGIAFYY
                                               A+DYR GS+F+++ S+  G+GIAFY+
                                               ALDYRFGSVFSSVVSICCGIGIAFYF
H_S09_0038098__  7740                          gcgtctgtgtatggtattgagagttt
                      GTACGGT  Intron 18  CAGctaagtgcttgcttctgggtgtctat
                      <0-----[7740 : 8372]-0>ttcctctgcctactcactcatcacttt
```



CAEEL-PGP-9____ 848 GWQMAFL VMAIFPFMAVGQALMMKYH
GWQMA+L +AIFP+ AVGQA++M++
GWQMALL TIAIFPLAAVGQAIQMRFM
H_S09_0038098__ 8451 gtcagct aagatctggggcgacaata
ggatcttGTGAGTT Intron 19 CAGctctttcctgactatgtt
agggata<0-----[8472 : 8531]-0>acctctgttggattggacg

CAEEL-PGP-9____ 874 GGSATSDAKEMENAGK TAMEAIENIR
G AT+DAKEMEN+GK AMEAIENIR
SGRATADAKEMENSGK IAMEAIENIR
H_S09_0038098__ 8589 tgcgagggagagaaga agaggagaac
cggcccacaataaggaGTGGGTC Intron 20 CAGtctactaatg
tgtcactggggactag<0-----[8637 : 8698]-0>tagaccgtta

CAEEL-PGP-9____ 900 TVQALTLQTKLYNIFCSHLDAPHGGNISKAIIR
TVQALTL+ +L+ FC HLD PH + KA+I+
TVQALTLERRLHAQFCHHLDGPHKTSRRKALIQ
H_S09_0038098__ 8729 agcgtatgccccgcttcccggccaaaacagcac
ctactctaggtacatgaatagcaacgggactta
gggagcgatgttcgctttactccaccaggtttg

CAEEL-PGP-9____ 933 GLTYGFANSIQFFTYAAAFRFGLFLI
G++YGFA+SI +F YA+ FRFGL+LI
GVSYGFASSIFYFLYASCFRFGLWLI
H_S09_0038098__ 8828 ggttgtgaaatttctgtttctgttca
GTGGGGA Intron 21 CAGgtcagtcggttattaccgtgtgtgtt
<0-----[8828 : 8907]-0>tcatattccctccgctatcccgggtc

CAEEL-PGP-9____ 959 FDKNVLMEPENVL 971
N + NVL
V--NGTIHSMNVL
H_S09_0038098__ 8986 g agaactaagca 9020
t agctactattg
c ccctcggttc

H_S09_0038098__ 9021 GTGAGCC Intron 22
<2-----[9021 :]-2>

Alternative splice

CAEEL-PGP-9____ 972 R
R
R
[agg]
H_S09_0038098__ Intron 22 CAG
<2-----[9021 : 9139]-2>

CAEEL-PGP-9____ 973 VLFAISFSFG
VLFAISF+ G
VLFAISFTAG
H_S09_0038098__ 9140 gctgattagg
tttctctccg
gaatattctca

CAEEL-PGP-9____ 983 TIGFAASYFPEYIKATFAAGLIFNMLEEEPRIDGMTSSGTYPQLSGEVK
++GFA+SYFPEYIKATFAAG+IF+MLEEEPRIDGMT++G P+++G VK
SMGFASSYFPEYIKATFAAGIIFHMLEEEPRIDGMTNNGKKPKITGAVK
H_S09_0038098__ 9171 aagtgatttcgtaagatgggaatcatgggccaggaaaagaacaaaggga
gtgtcgcatcaatacctccgtttattaaacgtagtcaagaacatcgcta
tgattcgcccgccagattctccccggagaaacccggtccgacgtctccg

CAEEL-PGP-9____ 1032 LNKVFFRYPERPAVPILQGLNVH VKP
LNKV+F+YPERP VPILQGL+++ VKP
LNKVYFKYPERPDVPILQGLDIN VKP
H_S09_0038098__ 9318 caagttatcgccggcaccgcgaa gac
taatataacagcatcttagtataGTAAGTC Intron 23 CAGtac
gtagtcgtcagacccataactcc<0-----[9387 : 9459]-0>aac

CAEEL-PGP-9____ 1058 GQTLALVGPSGCGKSTVISLLERLYDPLEGAV 1089
G+TLALVGPSGCGKSTVISLLERLYD L+G+V
GETLALVGPSGCGKSTVISLLERLYDALDGSV
H_S09_0038098__ 9469 ggacgcggcagtgatagatccgactggtggtg 9564
gactcttgcgggggaccttcttagtaactagct
cgttcgtgtttcagcataggcagccttgtttt

H_S09_0038098__ 9565 GTAAGTA Intron 24
<0-----[9565 :10861]-0>
10 n [9881 : 9890]

Alternative splice

CAEEL-PGP-9____ 972 R
R
R
[agg]
H_S09_0038098__ Intron 22 CAG
<2-----[9021 :10001]-2>
10 n [9881 : 9890]

CAEEL-PGP-9____ 973 VLFAISFSFG
VLFAISF+ G
VLFAISFTAG
H_S09_0038098__ 10002 gctgattagg
tttctctccg
gaatgttctta

CAEEL-PGP-9____ 983 TIGFAASYFPEYIKATFAAGLIFNMLEEEPRIDGMTSSGTYPQLSGEVK
++GFA+SYFPEYIKATFAAG+IF+MLEEEPRIDGMT++G P+++G VK
SMGFASSYFPEYIKATFAAGIIFHMLEEEPRIDGMTNNGKKPKITGAVK
H_S09_0038098__ 10033 aagtgatttcgtaagatgggaatcatgggccaggaaaagaacaaaggga
gtgtcgcatcaatacctccgtttattaaacgtagtcaagaacatcgcta
tgatctgtccgccagcctctccccggaaaaacccggtccgacgtctccg

CAEEL-PGP-9____ 1032 LNKVFFRYPERPAVPILQGLNVH VKP
LNKV+F+YPERP VPILQGL+++ VKP
LNKVYFKYPERPDVPILQGLDIN VKP
H_S09_0038098__ 10180 caagttatcgccggcaccgcgaa gac
taatataacagcatcttagtataGTAAGTA Intron 23 CAGtac
gtactcatcagactcataacccc<0-----[10249:10333]-0>gac

CAEEL-PGP-9____ 1058 GQTLALVGPSGCGKSTVISLLERLYDPLEGAV 1089
G+TLALVGPSGCGKSTVISLLERLYD L+G+V
GETLALVGPSGCGKSTVISLLERLYDALDGSV
H_S09_0038098__ 10343 ggacgcggcagtgatagatccgactggcggtg 10438
gactcttgcgggggaccttcttagtaactagct
cgttcgtgtctcagcatagacagtcttgcttt

H_S09_0038098__ 10439 GTAAGTA Intron 24
<0-----[10439:10861]-0>

H_S09_0038098__ Intron 24 CAG
<0-----[:10861]-0>

CAEEL-PGP-9____ 1090 TVDNNDLRQMNPKHLRKHIALVSQEP
+D NDLR++NP HLR HIALVSQEP
EIDGNDLREVNPTHLRAHIALVSQEP
H_S09_0038098__ 10862 gaggagtcggacactcgcagtgtcgc
atagaatgataccatgcatcttcaac
atttccatagcactgaccttgggagg

CAEEL-PGP-9____ 1116 ILFDTSIRENIVYGLQPGEYTHEQIETACSKANIHKFIDELPD
ILFD SIR+NI+YGL PG + + +ANIHKFI +LPD
ILFDRSIRDNILYGLPPGSVSDAAVHEVAQRANIHKFIMDLPD
H_S09_0038098__10940 actgatacgaactgcccgtgaggggcgggccgaacataaggcg
tttagctgaattagtccgctgacctaatcagcataatttatca
caccatctctcccctattccttgtgcacaatcctcaccgcgct

CAEEL-PGP-9____ 1159 GYETRVGEKGTQLSGGQKQRIAIARA
GY TR GEKGTQLSGGQKQRIAIARA
GYNTRAGEKGTQLSGGQKQRIAIARA
H_S09_0038098__11069 gtaacgggagacttggcaccagagcg
GTGAGTT Intron 25 CAGgaacgcgaagcatcggaaagtctcgc
<0-----[11069:11132]-0>tccgggaaaatagtacagaatccatt

CAEEL-PGP-9____ 1185 LIRNPKILLLDEATSALDTESEK QVQ
LIRNPK+LLLDEATSALDTESEK VQ
LIRNPKVLLLDEATSALDTESEK VVQ
H_S09_0038098__11211 caaacagccccgggaagtgagaga ggc
ttgacattttaaccgctacagaaGTTTCAT Intron 26 CAGtta
ccataatggtcgtctcgtcgcag<0-----[11280:11357]-0>tga

CAEEL-PGP-9____ 1211 VALDAAAKDRTCIVVAHRLSTIVNAGCIMVVKNGQVVEQ
ALD A++ RTCIVVAHRLST+VNA CIMVVK G+VVE+
EALDKASEGRTCIVVAHRLSTVVNANCIMVVKGGKVVEK
H_S09_0038098__11367 ggcgagtggcataggcccctaggagataaggaggaggga
actaaccaggcgtttcagtccttacagttttaggattaa
gctcgaaaccgtctgctgggtcctattagacgaaattaa

CAEEL-PGP-9____ 1250 G THNELIAKRGAYFALTQKQ 1269
G TH+EL+ +GAY+ALTQKQ
G THSELMQAKGAYWALTQKQ
H_S09_0038098__11484 g [gga] acagcacgaggttgcacac 11601
GTACAGA Intron 27 CAGgcagattacagcagctcaaa
<1-----[11485:11542]-1>acctaggacgcacggtagga

H_S09_0038098__11602 ACATTGG Intron 11852
<?-----[11602:]-?>

//

Gene 1
EXONS 1802 11601
Exon 1802 1900 phase 0
Exon 2007 2056 phase 0
Exon 2365 2482 phase 2

Exon 4236 4407 phase 0
Exon 4460 4577 phase 0
Exon 4715 4827 phase 1
Exon 4885 5001 phase 0
Exon 5283 5390 phase 0
Exon 5442 5537 phase 0
Exon 5599 5769 phase 0
Exon 5825 5914 phase 0

Exon 5976 6069 phase 0

Exon 6258 6351 phase 0

Exon 6914 6996 phase 1
Exon 7052 7090
7091 7148 phase 0
Exon 7209 7363 phase 1
Exon 7417 7524 phase 0 Exon 7578 7739 phase 0
Exon 8373 8471 phase 0
Exon 8532 8636 phase 0
Exon 8699 8827 phase 0
Exon 8908 9020 phase 0

Exon 9140 9386 phase 2
Exon 9460 9564 phase 0

Exon 10002 10248 phase 2
Exon 10334 10438 phase 0

Exon 10862 11068 phase 0
Exon 11133 11279 phase 0
Exon 11358 11484 phase 0
Exon 11543 11601 phase 1
//

Making a N in phase 2 intron
Making a E in phase 1 intron

Making a E in phase 1 intron

Making a K in phase 1 intron

Making a E in phase 1 intron

Making a R in phase 2 intron

Making a R in phase 2 intron

Making a G in phase 1 intron

>H_S09_0038098__HAECO_S09_Supercontig_0038098
_1802-2482bp_AA

_4236-5914bp_AA
_5976-6069bp_AA 6258-6351bp_AA_Alternative_splice
_6914-7090bp_AA
_7091-9020bp_AA
_9140-9564bp_AA 10002-10438bp_AA_Alternative_splice
_10862-11601bp_AA
_CAEEL-PGP-9_35-123AA---268-596AA_597-627AA_Alternative_splices_628-661AA_662-971AA_972-1089AA_Alternative_splices_1090-1269AA

FRYTTTFDKVLLMIGSIVAIGTGIGLPMM
SIIMGNISQNFMNINGNTTSNLKKLLLLSSAINQFEHDVIQNCLKYVYLGCGIFTAATIQ

YQKALED
GKSTGIKKSFYIGVGLGITFLIMFSSYCLAFWVGTDFVFKGQMNGGTVMTVFFSVMMGSM
ALGQAGPQFAVLGTAMGAAGSLYQIIDRVEIDSYSTDGVKPSNLKGKVTVSNLKFTYPTR
PDVPILK
GVSFEANPGETIALVGSSGCGKSTIIQLLLRYYN
PEDGKITIDGAEIDKINIEFLRNYVGVVSQEPMLFNTTIEQNIRYGRENVTDAEITAALR
KANAYNFVQSFPDGIYTNVGDRGTQMSGGQKQRIAIARALVRDPKILLLDEATSALDAES
EHVVQQALENASKGRTTIVVAHRLSTIRNADRIIAIKDGE

VMEVGTHDELMARKGLYHELVNSQVFVDVDGE

VMEVGTHDELIARKGLYHELVNAQVFADVDDK

SGEPGERRRTMSSSRSRSPSLASPEYKRLKSQMSTEIAAG

GGAQNDPV
KAEKDLERLKKELEEEGAAKANLFKILGYARPEWPFIALAVTSSIVQGCVFPAFSLFFSQ
IID
VFSKQPGDPTLKSDGHFWALMFLVLGGTQAMTMLIQCFFFGLSAERLTMRLRSKIF
QNVMRMDATYFDMPRHSAGKITTRLATDAPNVKSALDYRFGSVFSSVVSICCGIGIAFYF
GWQMALLTIAIFPLAAVGQAIQMRFMSGRATADAKEMENSGKIAMEAIENIRTVQALTLE
RRLHAQFCHHLDGPHKTSRRKALIQGVSYGFASSIFYFLYASCFRFGLWLIVNGTIHSMN
VL

```
RVLFAISFTAGSMGFASSYFPEYIKATFAAGIIFHMLEEEPRIDGMTNNGKKPKITGA
VKLNKVYFKYPERPDVPILQGLDINVKPGETLALVGPSGCGKSTVISLLERLYDALDGSV

RVLFAISFTAG
SMGFASSYFPEYIKATFAAGIIFHMLEEEPRIDGMTNNGKKPKITGAVK
LNKVYFKYPERPDVPILQGLDINVKP
GETLALVGPSGCGKSTVISLLERLYDALDGSV

EIDGNDLREVNPTHLRAHIALVSQEPILFDRSIRDNILYGLPPGSVSDAAVHEVAQRANI
HKFIMDLPDGYNTRAGEKGTQLSGGQKQRIAIARALIRNPKVLLLDEATSALDTESEKVV
QEALDKASEGRTCIVVAHRLSTVVNANCIMVVKGGKVVEKGTHSELMQAKGAYWALTQKQ
//

>H_S09_0038098___HAECO_S09_Supercontig_0038098
_1802-2482bp
---
_4236-5914bp
_Alternative_splice_5976-6069bp_6258-6351bp
_6914-7090bp
_7091-9020bp
_Alternative_splice_9140-9564bp_10002-10438bp
_10862-11601bp

ttccgctacacgactacatttgataag
gtgctgctcatgattggctcaattgtcgctattggtactggtataggacttccaatgatg
tctattattatgggtaatatttcacaaaacttcatgaatatcaatggaaacactactagt
aatttaaagaagcttttacttttatcttcagcgatcaatcaatttgaacatgatgtgatc
cagaattgtttgaaatacgtctatctcggctgtgggatattcacggctgccacaatacag

gtaccaaaaagctttggaagacggtaaatcgacgggaatcaagaaatctttttatatt
ggcgttggtctgggaataacgttcctgatcatgttctcctcgtattgcctggcattttgg
gtcggcacggactttgtcttcaaaggtcaaatgaatggaggaactgtgatgacagttttc
ttctccgtaatgatgggttccatggcgctcggacaagctggacctcagtttgctgtcctc
ggtacggctatgggtgcagccggatctctctatcagattattgatcgggtagaaatcgat
tcctactcaacggatggcgtgaaaccatcgaatctcaaagggaaagtaacagtctcaaat
ctgaaattcacttatccaacgcgaccagatgttcctattcttaag

ggagtctcattcgaagca
aatcctggtgaaacgatagcattggtgggttccagtggttgtggaaaaagtaccataatc
cagttgttactgcgatattacaatcctgaagatggaaagatcacgatagatggtgcggaa
atcgacaagattaacattgagttccttcgaaactacgtcggagtcgtctcacaagagccg
atgcttttcaatacaaccattgaacagaacatccggtacggacgagagaatgtcaccgat
gcggaaatcaccgctgctctacgtaaagcaaacgcttacaactttgtgcaatcgtttcct
gatggtatctacacgaatgtcggtgaccgtggtacccagatgtccggtggccagaagcag
cgtatcgccattgctcgtgcgttggtcagagatccgaaaattctcttgctggatgaggcc
actagcgctctcgacgccgaaagtgaacatgtcgtccagcaggctctggaaaacgcttct
aaaggaagaactaccatcgtcgtagcccatcgattgtcgactattcgaaatgctgacaga
atcattgccattaaagatggagaa

gtgatggaagtcggcactcatgacgagctgatggcccggaaaggactgtaccatgaattg
gtcaattctcaggtgttcgtcgatgtcgacgggg

gttatggaagtcggcacacatgatgagctgatcgca
cgcaagggactctaccatgagctggtcaatgcccaggtgttcgccgacgtggacgaca

aa
agtggcgagccgggtgaacgacggcgtacgatgtcatcttcacgttcgagatctccatcg
ttagcttcacccgaatataagcggctcaaatctcaaatgtcaactgaaattgctgctggt

ggaggtgctcagaacgatccagtgaaagcagagaaagacttggaacgactgaagaaagaa
ctggaagaggaaggtgccgccaaagcgaatcttttcaaaatccttggctacgctcgacct
gaatggccttttatcgccctcgccgttacctcttcaatcgttcaaggctgtgtttttcca
gcattttcactctttttctcacaaatcatcgat
gtgttttccaag
cagccaggtgatccgacgctcaagagtgacggccatttctgggcacttatgttcctcgtg
ctgggtggcactcaggcgatgaccatgctcattcag
tgcttcttcttcggtttgtcagcc
gaacggctcacaatgcgattacgatcaaaaattttccaaaatgtgatgagaatggatgcc
acctacttcgacatgcctcgccattcagctggaaaaatcactacccgcttggcgacagat
gcaccaaatgtgaaatcggctcttgactaccgtttcggttcggtcttcagttcagtcgtt
tccatatgctgtggcataggtatcgcattctattttggatggcagatggcacttttaaca
atcgccatttttcccttggctgctgtggggcaagctattcagatgagattcatgtctggg
cgtgccacagccgatgcgaaggagatggaaaacagtggaaagattgcaatggaagccatc
gagaatattcgaacggtgcaggcattgaccttggaacgtcggcttcatgcccagttctgt
catcatctagacggtcccccacaaaaccagcagacggaaggctcttattcagggtgtctca
tatggatttgctagcagcatcttttacttcctgtacgcttcatgtttccgcttcgggttg
tggcttatcgtcaacggcaccattcactcgatgaatgttctcag

ggtactatttgcaatttctttcactgccggaagtatgggatttgctagctcgtacttcccc
gagtacatcaaagcgacatttgctgccggtatcatcttccacatgttggaagaggaacca
cgaatcgacggcatgacgaataacggcaagaaacccaagattaccggtgccgtcaagctg
aataaagtgtatttcaagtatcccgaacggccagacgtccccatacttcaaggactcgat
atcaacgtaaaacccggcgagactcttgccctggttgggcctagtggttgcggaaagtcc
acagttatatcgctgctcgaaaggctctacgatgctttggatggttctgtt

ggtactatttgcgatt
tctttcactgct
ggaagtatgggatttgccagttcgtatttccccgagtacatcaaagcg
accttcgctgccggtatcatcttccacatgttggaagaagaaccacgaatcgacggcatg
acgaataacggcaagaaacccaagattaccggtgccgtcaagctgaataaagtctatttc
aaatatcccgaacggccagacgttcccatacttcaaggactcgacatcaacgtgaaaccc
ggcgagactcttgccctggttgggcctagcggttgcggaaagtccacagttatatcgcta
ctcgaaaggctttacgatgctctggacggttctgtt

gaaattgatggtaacgacttacgt
gaagtgaacccaacccatttgcgagcccacattgctttggtgtcgcaagagccgatccta
ttcgacagatctatccgtgacaatatcctctacggccttccacctggttccgtcagtgat
gcggctgtgcacgaagtcgcacaacgtgccaacattcacaaattcatcatggacctgccc
gatggttacaacacgcgggcgggagaaaaaggaactcaattgtctggaggccaaaagcaa
cgaattgccatcgcacgtgctctcatcagaaatccaaaagttctgctgcttgacgaggct
accagtgccttggatacgagagcgaaaaggttgtgcaagaggcccttgacaaggcatca
gaaggccgcacgtgtatcgttgtggcccatcggctgtcgactgtcgtcaatgcaaattgt
ataatggtagtcaagggaggaaaagttgttgaaaaaggaacccacagtgaactgatgcaa
gccaagggcgcatactgggcgcttacacagaagcaa
//
```

A.25 *HAECO*_S09_SUPERCONTIG_0046372 ≈ *CAEEL*-PGP-9

```
genewise
Query protein:       CAEEL-PGP-9____CAEEL-PGP-9 C47A10 1 1-1294AA
Target Sequence      H_S09_0046372___HAECO_S09_Supercontig_0046372

H_S09_0046372__  171           Intron 1   CAG                          365
                  <1-----[171  :  365]-1>


CAEEL-PGP-9____   94  RAEFSHEVIQNCLKYVYLGCG
                        F H+V QNCLKYVYLGCG
                      FQQFEHDVTQNCLKYVYLGCG
H_S09_0046372__  366  tcctgcggacattatgtcgtg
                    ctaataaatcaagtaatatggg
                    tcaataccgcactgacccctta


CAEEL-PGP-9____  116 IFAAGFLQ                        ASCFMVICEKLSNRFRRQ
                     IFAA  +Q                        A+CF+ + E L N++RRQ
                     IFAAATIQ                        ATCFLTVGENLVNQLRRQ
H_S09_0046372__  431 atgggaac                        gattcagggacgaccacc
                     ttccccctaGTATAGC  Intron 2   AAGccgttctgaattaatgga
                     acgaagtg<0-----[455  : 1391]-0>gactgttagtgataggag


CAEEL-PGP-9____  142 FFHSVMRQEIAWYDKNTSGTLSNKLF              D
                     FF S++RQ+I W+DKN SGTL+ KLF              D
                     FFKSILRQDIPWFDKNGSGTLATKLF              D
H_S09_0046372__ 1446 ttatacccgacttgaagtgatgaattg          [gac]
                     ttacttgaatcgtaaagcgctccattaGTAAGTG  Intron 3
                     ccaattcactagttgttaaagcaggt <2-----[1526 : 1584]


CAEEL-PGP-9____  168     NLERVREGTGDKVGLAFQMMAQFIGGFAVAFTYDWLLTLIMMSLS
                         NLERV+EGTGDKVGL +Q +AQF+GGF VAFTYDW LTLIMMSL+
                         NLERVKEGTGDKVGLMIQYVAQFFGGFIVAFTYDWKLTLIMMSLA
H_S09_0046372__ 1582     acgcgaggaggaggcaactggcttggtaggtatgtacacaaatcg
                    AAG atagtaagcgaatgtttaatcattggtttctcaagatcttttctc
                    -2>ctcaacgaacccactggcatgagcctcccgacctcgacggtggtgt


CAEEL-PGP-9____  214 PFMMICGLFLAK                                  225
                     PFM++CG F+A+
                     PFMIMCGAFIAR
H_S09_0046372__ 1721 ctaaatggtaga                                 1756
                     cttttggcttcgt
                     acgcgcaatttga


H_S09_0046372__ 1757 GTAGGCTA  Intron                              1847
                     <0-----[1757 :     ]-0>
//

Gene 1
EXONS 366 1756


  Exon 366 454 phase 1
  Exon 1392 1525 phase 0
  Exon 1585 1756 phase 2
//

Making a D in phase 2 intron
```

```
>H_S09_0046372___HAECO_S09_Supercontig_0046372
_366-1756bp_AA
_CAEEL-PGP-9_94-225AA

FQQFEHDVTQNCLKYVYLGCGIFAAATIQA
TCFLTVGENLVNQLRRQFFKSILRQDIPWFDKNGSGTLATKLFDNLERVKEGTGDKVGLM
IQYVAQFFGGFIVAFTYDWKLTLIMMSLAPFMIMCGAFIAR
//

>H_S09_0046372___HAECO_S09_Supercontig_0046372

ctttccaacaatttgaacacgacgtgacccaaaactgtttgaaatac
gtctacctcggttgtggaatattcgcggcagcaacgattcaggcgacatgctttctgact
gttggagagaatctggtaaatcaactgaggcgacagttcttcaaatcaattcttcgccaa
gacattccatggtttgataagaatggttcaggaacattggccacaaagttgtttgacaat
ctcgaacgagtcaaggaaggaaccggcgacaaagtcggtctgatgatccaatatgtggca
cagttcttcggtggcttcatcgtggcattcacctatgactggaaaactcacgctgattatg
atgtctctggctccattcatgatcatgtgcggagcatttattgctagg
//
genewise
Query protein:       CAEEL-PGP-9____CAEEL-PGP-9_C47A10_1_1-1294AA
Target Sequence      H_S09_0046372___HAECO_S09_Supercontig_0046372

H_S09_0046372__ -365 [365  :  171]                                  171


CAEEL-PGP-9____  676 KRLKKELEEEGAVKANLFKILRYARPEWIYIFFAIIAALIQGAVMPAFS
                     ++ K++LEEEGAV+ANL+KI RYARPEW +I +A++++++QG V+PAFS
                     EKQKEKLEEEGAVRANLIKIFRYARPEWPFITIAVLSSIVQGCVFPAFS
H_S09_0046372__ -170 gacagacgggggggagacaaatatgccgtctaaaggcttagcgtgtcgtt
                     aaaaaataaagctgcattattgacgcagcttctcttccttaggttcctc
                     agaaaacaaattcagcctactgttacagttcttctccacaaacttatcc


CAEEL-PGP-9____  725 LFFSQII                                        731
                     LFF+QII
                     LFFTQII
H_S09_0046372__  -23 tttacaa                                         3
                     tttcatt
                     accaatc


H_S09_0046372__   -2 GA                                              1
                     [2    :    1]
//

Gene 1
EXONS 170 3
  Exon 170 3 phase 0
//

>H_S09_0046372___HAECO_S09_Supercontig_0046372
_170-3bp_AA
_CAEEL-PGP-9_676-731AA

EKQKEKLEEEGAVRANLIKIFRYARPEWPFITIAVLSSIVQGCVFPAFSLFFTQII
//

>H_S09_0046372___HAECO_S09_Supercontig_0046372
gaaaagcaaaaagaaaaactcgaagaagaaggtgctgtcagagcgaacctcattaaaatc
tttaggtatgctcgacccgaatggccttttatcactattgccgttctctcctcaatcgta
caaggatgcgtttttccagctttctccttattcttcacacaaattatc
//
```

A.26 *HAECO*_S09_SUPERCONTIG_0016882 ≈ *CAEEL*-PGP-9

```
genewise
Query protein:       CAEEL-PGP-9____CAEEL-PGP-9 C47A10 1 1-1294AA
Target Sequence      H_S09_0016882___HAECO_S09_Supercontig_0016882

H_S09_0016882__     1          Intron     TAG                          97
                    <0-----[     :   97]-0>


CAEEL-PGP-9____  1090 TVDNNDLRQMNPKHLRKHIALVSQEPILFDTSIRENIVYGLQPGEYTHEQ
                       VD NDLR+MNP HLR HIALVSQEPILFD SIR+NI+YGL PG  +  +
                      EVDGNDLREMNPTHLRAHIALVSQEPILFDRSIRDNILYGLPPGSVSEAE
H_S09_0016882__    98 ggggagtcgaacacccgcagtgtcgcactgatacgaactgcccgtgaggg
                      atagaatgataccatgcatcttcaactttagctgaattagtccgctgaca
                      acttccacagcttcgtcctagaaggacctcaccgcccctcgaatcctata


CAEEL-PGP-9____  1140 IETACSKANIHKFIDELPD                        GYETRVG
                      +     +ANIHKF+ ELP+                        GY TR G
                      VHEVAQRANIHKFVMELPE                        GYNTRAG
H_S09_0016882__   248 gcgggccgaacatgagtcg                        gtaacgg
                      taatcagcataatttatcaGTGCGTT  Intron 1   TAGgaacgcg
                      gtacaattcccattgagcg<0-----[305   :   368]-0>cccatga


CAEEL-PGP-9____  1166 EKGTQLSGGQKQRIAIARALIRNPKILLLDEATSALDTESEK
                      EKG QLSGGQKQRIAIARALIRNPKILLLDEATSALDTESEK
                      EKGVQLSGGQKQRIAIARALIRNPKILLLDEATSALDTESEK
H_S09_0016882__   390 gaggcttggcaccagagcgcaaacaatccgggaagtgagaga
                      aagtatcggaaagtctcgcttgacattttaaccgctacagaa
                      agccggttggaagcccattccatgatagtcgtcttgtcacag


CAEEL-PGP-9____  1208                        QVQVALDAAAKDRTCIVVAHRLSTIV
                                              VQ ALD A++ RTCIVVAHRLST+V
                                             VVQEALDKASEGRTCIVVAHRLSTVV
H_S09_0016882__   516                        ggcggcgagtggcatagggccctagg
                      GTGCTAT  Intron 2   CAGttaactaaccaggcgtttcagtcctt
                      <0-----[516   :   594]-0>tgaggccgaaaccctctgatgggttc


CAEEL-PGP-9____  1234 NAGCIMVVKNGQVVEQ              G          THNELIA
                      NA CIMVVK G+VVE+              G          TH+EL+
                      NANCIMVVKGGKVVEK              G          THSELMQ
H_S09_0016882__   673 agataaggaggagggag          [gga]         acagcac
                      acagttttaggattaa GTACAGC  Intron 3   CAGgcagatta
                      tcttagacgaaataaa <1-----[722   :   781]-1>acttagga


CAEEL-PGP-9____  1258 KRGAYFALTQKQS
                       +GAY+ALTQKQ+
                      AKGAYWALTQKQT
H_S09_0016882__   805 gaggttgcacaca
                      cagcagctcaaac
                      cgtacggtaggaa

CAEEL-PGP-9____  1271 SNQSGGAFDTSEALDDDDDDHVKF                      1294
```

```
                      LAKG--------------------
H_S09_0016882__   844 tgag                                          855
                      tcag
                      gcga


CAEEL-PGP-9____       *
                      *
                      *
H_S09_0016882__   856 t                                             858
                      g
                      a


H_S09_0016882__   859  [859  : 1712]                                1712
//

Gene 1
EXONS 98 855
  Exon 98 304 phase 0
  Exon 369 515 phase 0
  Exon 595 721 phase 0
  Exon 782 855 phase 1
//

Making a G in phase 1 intron


>H_S09_0016882___HAECO_S09_Supercontig_0016882
_98-855bp_AA
_CAEEL-PGP-9_1090-1294AA_End

E
VDGNDLREMNPTHLRAHIALVSQEPILFDRSIRDNILYGLPPGSVSEAEVHEVAQRANIH
KFVMELPEGYNTRAGEKGVQLSGGQKQRIAIARALIRNPKILLLDEATSALDTESEKVVQ
EALDKASEGRTCIVVAHRLSTVVNANCIMVVKGGKVVEKGTHSELMQAKGAYWALTQKQT
LAKG
//

>H_S09_0016882___HAECO_S09_Supercontig_0016882
_98-855bp

gaa
gtcgatggtaacgacttacgcgaaatgaaccctactcacctgcgtgcccacattgcattg
gtatcacaggagccaatcctctttgacagatccatccgggacaacatcctctatggcctg
ccaccaggttccgtcagtgaagctgaagtgcatgaagtcgcacaacgtgctaacatccac
aaatttgttatggaattgcccgagggctacaacacacgtgcgggagaaaagggcgtccag
ttgtctggtgggcagaaacaacggatcgccatcgcacgtgctctcatcagaaatccgaaa
attttactgcttgacgaggctaccagtgctttggataccgaaagcgaaaaggttgtgcaa
gaggcgctcgacaaggcatcagaaggccgcacctgtatcgttgtggcacatcggctgtcg
actgttgtcaatgccaattgtataatggtagtcaagggaggaaaagttgtagaaaaagga
acccatagtgaactgatgcaagccaagggtgcatactgggcgcttacacagaagcaaaca
ttggccaagga
//
```

A.27 *HAECO*_S09_SUPERCONTIG_0035472 ≈ *CAEEL*-PGP-10

```
genewise
Query protein:      CAEEL-PGP-10___CAEEL-PGP-10 C54D1 1 1-1382AA
Target Sequence     H_S09_0035472___HAECO_S09_Supercontig_0035472

H_S09_0035472__-1063
                     <?-----[1063 :  397]-?>


CAEEL-PGP-10___  120 NFVSISILREEGKLGLKNLTSDYPIDDQFAHSATPAFISMLGLSIAMFI
                     +F  +   +EE    +    +   +D  FA SATPAFI ML LS++MF
                     SFAKMLPFKEERLYNISGAANGSKFDADFAASATPAFIIMLSLSVSMFS
H_S09_0035472__ -397 atgaatctaggcctaatgggagtatgggtggtgacgtaaattttgaatt
                     gtcattctaaagtaatcgccagcatacatccccccttttctctgttc
                     tttagaccgaacgtttgttgtaaactgtcgaaaaacttagaaatgcgtt


CAEEL-PGP-10___  169 AAFCQRIAWEI                                    179
                     AAF Q   +EI
                     AAFIQVNFFEI
H_S09_0035472__ -250 ggtacgattgat                                    217
                     ccttatattat
                     tattggctcat
```

```
H_S09_0035472__ -216 GTGTTTT  Intron                                  1
                     <1-----[216  :     ]-1>
//

Gene 1
EXONS 397 217
  Exon 397 217 phase 0
//

>H_S09_0035472___HAECO_S09_Supercontig_0035472
_397-217bp_AA
_CAEEL-PGP-10_120-179AA

SFAKMLPFKEERLYNISGAANGSKFDADFAASATPAFIIMLSLSVSMFSAAFIQVNFFEI
//

>H_S09_0035472___HAECO_S09_Supercontig_0035472

agttttgctaaaatgttacccttcaaggaagaacgcctgtataatatttcgggtgctgcg
aatggatcaaaattcgatgcggatttcgcggcatcagcaacaccagcctttattataatg
ttatcattatctgtgagcatgttttctgctgcatttattcaggtgaacttttcgaaatt
t
//
```

A.28 *HAECO*_S09_SUPERCONTIG_0025718 ≈ *CAEEL*-PGP-10

```
genewise
Query protein:      CAEEL-PGP-10___CAEEL-PGP-10 C54D1 1 1-1382AA
Target Sequence     H_S09_0025718___HAECO_S09_Supercontig_0025718

H_S09_0025718_-13784          Intron   CAG
                      <2-----[     : 2963]-2>
                        10 n [6325 : 6316]
                       307 n [12361:12055]


CAEEL-PGP-10___   219  SADSIYNGISDHLPMVVFILSYLLVNI            G
                         D+IYNGISD++PMV+FI +YL+ NI
                       YTDTIYNGISDNIPMVIFISAYLVANI            A
H_S09_0025718__-2964  tagaatagatgaacagatatgtcggaag          [gct]
                       acactaagtcaatctttttccattcat GTGTGTT  Intron 1
                     atactactgccccaagcactcgttctta <1-----[2879 : 2019]


CAEEL-PGP-10___   246      VCLYIQWDVTLFMCLAIPLLIISRILYSK
                           VCLYIQWDVTL+MC AIP+LIISRI++SK
                           VCLYIQWDVTLLMCSAIPMLIISRIVFSK
H_S09_0025718__-2021       gtttactggactattgacacaaacagtaa
                     CAGctgtatagatctttgcctcttttggtttga
                     -1>tacataagtgctagcaatagctacctgttg


CAEEL-PGP-10___   276                          WFANTMEEETKLQNKITNLVNETFNC
                                               WF+ TM++E  LQNKI+NLVNETF+C
                                               WFSKTMDQEVHLQNKISNLVNETFSC
H_S09_0025718__-1929                           tttaaagcggcccaaaaacgagatat
                     GTCAGAC  Intron 2   CAGgtcactaaatataaatgattaactgg
                     <0-----[1929 : 1853]-0>gtggcgtaatccacatctagtaaccc


CAEEL-PGP-10___   302 ITTVISFAAQKQKIN              K          FEKLSAEH
                      I TVISFAAQKQ I                K          FE+LS E+
                      IRTVISFAAQKQTIT               K          FERLSMEN
H_S09_0025718__-1774  acagaatggcacaaaa           [aaa]         tgactaga
                      tgcttgtccaaactcaGTTAGTT  Intron 3   GAG tagtctaa
                      aatactcttagagac <2-----[1727 : 1660]-2>atagacggt


CAEEL-PGP-10___   326 SKLTEERLRSSVVFDSLTQ                       ILLTELI
                      +KLTE RLRSS V+DSLTQ                       ILLTELI
                      NKLTESRLRSSTVYDSLTQ                       ILLTELI
H_S09_0025718__-1634  aacagtcccttagtgacac                       accagta
                      aatcacgtgccctaagtcaGTAATAA  Intron 4   CAGtttcatt
                      tacaggcgtcaagtttgag<0-----[1577 : 1498]-0>ctacagt
```

```
CAEEL-PGP-10___   352 FTGALCYGIWRVADNSPGRLCA-LAIN                    377
                       FT ALCYGIWRVAD+SPGRL A LA++
                       FTAALCYGIWRVADHSPGRLAAVLAFD
H_S09_0025718__-1476  taggtttgatagggctcgatgggcgtgt                  1394
                       tccctgagtggtcaaccggtccttctat
                       ttatgtcaagagcttataaaagaattt


H_S09_0025718__-1393  GTTTTAC  Intron                                  1
                       <2-----[1393 :     ]-2>
//

Gene 1
EXONS 2964 1394
  Exon 2964 2880 phase 0
  Exon 2018 1930 phase 1
  Exon 1852 1728 phase 0
  Exon 1659 1578 phase 2
  Exon 1497 1394 phase 0
//

Making a A in phase 1 intron
Making a K in phase 2 intron


>H_S09_0025718___HAECO_S09_Supercontig_0025718
_2964-1394bp_AA
_CAEEL-PGP-10_219-377AA

YT
DTIYNGISDNIPMVIFISAYLVANIAVCLYIQWDVTLLMCSAIPMLIISRIVFSKWFSKT
MDQEVHLQNKISNLVNETFSCIRTVISFAAQKQTITKFERLSMENNKLTESRLRSSTVYD
SLTQILLTELIFTAALCYGIWRVADHSPGRLAAVLAFD
//

>H_S09_0025718___HAECO_S09_Supercontig_0025718

atataca
gacactatatacaatgggatctccgacaacataccaatggtcatattcatttccgcgtat
cttgtcgctaatatagctgtatgcttatatatacaatgggatgtgacccttttaatgtgc
tcagcaattccaatgctcattataagccgcattgtgtttagtaagtggttttcgaagacc
atggatcaagaagttcacctccaaaacaaaattagcaatctagtgaatgaaacattcagc
tgcatacgaactgtaatcagtttcgctgctcaaaagcaaacgataaccaaatttgaaagg
ctatccatggagaataataaactcacagagtcgcgcctgcgttcctcaacagtgtatgat
agtctgacacagatccttctaaccgaattgatttttactgcagctttgtgttacggaata
tggagagtggccgatcattcacctggaagattagcagcggtactagcttttgattt
//
```

A.29 *HAECO*_S09_SUPERCONTIG_0064200 ≈ *CAEEL*-PGP-10

```
genewise
Query protein:      CAEEL-PGP-10___CAEEL-PGP-10 C54D1 1 1-1382AA
Target Sequence     H_S09_0064200___HAECO_S09_Supercontig_0064200

H_S09_0064200__    1          Intron    CAG                          23
                   <0-----[     :   23]-0>


CAEEL-PGP-10___  374 LAINMLYMCVTSISIGFHMNGATR                        397
                     LAINMLYMCVTSISIGFH+NGA+
                     LAINMLYMCVTSISIGFHINGAST
H_S09_0064200__   24 cgaaactatgatatagtcaaggta                         95
                     tctattatgtcctctgtatagccc
                     ccccgctgcaaactcgttctagga


CAEEL-PGP-10___  398 AK                                               399
                     x!!!!
H_S09_0064200__   96 gtttt                                           102
                     N
                     c


CAEEL-PGP-10___  400 DNAKEICEVL
                      +A ++  +L
                     QSASQLRAIL
H_S09_0064200__  103 cagtcccgac                                      132
                     agccatgctt
                     acataaataa


CAEEL-PGP-10___  410                N          EQPKIEVELASI-DNNIDYVPVT
                                               E P+IE +     DN    +P
                                   D           ECPRIESDYGFAGDNPTRCLPAP
H_S09_0064200__  133 g           [gat]         gtccagagtgtgggacacttcgc
                      GTCAAAC  Intron 1   TAGaagcgtagaagtcgaaccggtccc
                      <1-----[134  :  195]-1>tacgatattttcttctcaacgaca


CAEEL-PGP-10___  433 RCKYRRQSMKFM              G          KGALHFRDIHF
                       KYRRQSMK+M              G          KGA+HFRDIHF
                     --KYRRQSMKYM              G          KGAIHFRDIHF
H_S09_0064200__  267   atcccaaatag           [gga]         aggactcgact
                       aaggagtaat GTATAGA  Intron 2   AAGgagctatgatat
                       gccgacgatg <1-----[298  :  380]-1>aatccttatccc


CAEEL-PGP-10___  457 SYPSRKETE                        VLRGISFKVEAGEKIAI
                     SYPSR + E                        VL+G++F VEAGEKIA+
                     SYPSRPDVE                        VLKGVTFHVEAGEKIAL
H_S09_0064200__  416 atcaacggg                        gtaggatcggggaagc
                     gacggcataGTAAATG  Intron 3   TAGttagtctatacgaatct
                     ctacgaccg<0-----[443  :  585]-0>tggatgttcattggcag
```

```
CAEEL-PGP-10___  483 VGSSGSGKSTLTALLLRFYDPTKGA                       507
                     VGSSGSGKSTLTALLLRFYDP  G+
                     VGSSGSGKSTLTALLLRFYDPDSGS
H_S09_0064200__  637 ggtagtgaaacagctccttgcgagt                       711
                     tgcggcgagctcctttgtaacaggc
                     tcttcccacttaatacacccttcag


H_S09_0064200__  712 GTAAGTG   Intron                                 1168
                     <0-----[712  :     ]-0>
//

Gene 1
EXONS 24 711
  Exon 24 95
       96 102
       103 133 phase 0
  Exon 196 297 phase 1
  Exon 381 442 phase 1
  Exon 586 711 phase 0
//

Making a D in phase 1 intron
Making a G in phase 1 intron


>H_S09_0064200___HAECO_S09_Supercontig_0064200
_24-711bp_AA
_CAEEL-PGP-10_374-397AA_XX_400-507AA

LAINMLYMCVTSISIGFHINGAST

XX

QSASQLRAILDECPRIESDYGFAGDNPTRCLPAP
KYRRQSMKYMGKGAIHFRDIHFSYPSRPDVEVLKGVTFHVEAGEKIALVGSSGSGKSTLT
ALLLRFYDPDSGS
//

>H_S09_0064200___HAECO_S09_Supercontig_0064200

ctcgccatcaacatgctctatatgtgcgtaacatcaatctctatcgggtttcatatcaat
ggagcgtcgaca

gnc
tttt

caaagcgcatctcaactacgagctatactagatgaatgcccgcgaattgaaagtgattat
ggtttcgctggtgacaatcccacacgatgcttgccagccccaaagtaccgccggcaaagc
atgaaatatatgggaaaaggtgccatccatttttcgagatatccacttcagctatccaagc
aggccagacgtcgaggttttgaagggagttacgtttcatgtcgaagctggtgagaagatc
gcactggttggctctagtggctccggcaaaagcactcttacagcacttttactccgattc
tacgaccctgatagcggatcg
//
```

A.30 *HAECO*_S09_SUPERCONTIG_0038139 ≈ *CAEEL*-PGP-10

```
genewise
Query protein:      CAEEL-PGP-10___CAEEL-PGP-10 C54D1 1 1-1382AA
Target Sequence     H_S09_0038139___HAECO_S09_Supercontig_0038139

H_S09_0038139__-7037                    CAG
                    <0-----[     : 5392]-0>


CAEEL-PGP-10___   508 ILLDGENIKTMCPDDLRGQCSLVSQEPVLFDGTISDNIRYGRLDATQ
                        ILLDG+N+KT+CPDDLRG CSLVSQEPVLFDGTISDNIRYGRLDATQ
                        ILLDGDNLKTLCPDDLRGMCSLVSQEPVLFDGTISDNIRYGRLDATQ
H_S09_0038139__-5391    accgggataattcggtcgatttgtcgcgctggaaagaactgccggac
                        tttagaatactgcaatggtgcttcaactttagctgaatgaggtacca
                        tgcctttaaaacaccatagcaatcggcggccaccttcaaccatccag


CAEEL-PGP-10___   555                         QEINDAARKVGAWKFINSLPDGMQTR
                                              QEINDAARKVGAW+FI+SLP+GMQ
                                              QEINDAARKVGAWQFISSLPEGMQxx
H_S09_0038139__-5250                          cgaagggcagggtctaattcggacNN
                    GTCGAGC  Intron 1   CAGaataaccgatgcgattgctcagtaNN
                    <0-----[5250 : 4858]-0>aacctttaagaggaccttacaagaNN


CAEEL-PGP-10___   555                         QEINDAARKVGAWKFINSLPD
                                              QEINDAARKVGAW+FI+SLP+
                                              QEINDAARKVGAWQFISSLPE
H_S09_0038139__-5250                          cgaagggcagggtctaattcg
                    GTCGAGC  Intron 1   CAGaataaccgatgcgattgctca
                    <0-----[5250 : 4858]-0>aacctttaagaggaccttaca


CAEEL-PGP-10___   576 GMQ                                        578
                        GMQ
                        GMQ
H_S09_0038139__-4794    gac                                    4786
                        gta
                        aga


CAEEL-PGP-10___   579 TRVG                                       582
```

```
                          +
                        xxxx
H_S09_0038139__-4785    NNNN                                   4774
                        NNNg
                        NNNa
                              10 n [4785 : 4776]
                            1404 n [3076 : 1673]


H_S09_0038139__-4773                  [4773 :    1]                1
//

Gene 1
EXONS 5391 4786
  Exon 5391 5251 phase 0

  Exon 4857 4786
       4785 4774 phase 0
//

>H_S09_0038139___HAECO_S09_Supercontig_0038139
_5391-4786bp_AA
_CAEEL-PGP-10_508-578AA

ILLDGDNLKTLCPDDLRGMCSLVSQEPVLFDGTISDNIRYGRLDATQQEINDAARKVGAW
QFISSLPEGMQ

xxxx
//

>H_S09_0038139___HAECO_S09_Supercontig_0038139

attctgctcgacggtgataatttaaaaacattatgcccagacgacttacgtggaatgtgc
tcattagtttcccaggagcccgtgctgttcgacggaaccatcagtgataacatacgatac
ggccgacttgacgccacacagcaagaaatcaacgatgctgctcgaaaagtgggagcgtgg
caattcatcagttctttacccgaaggaatgcaa

nnnnnnnnnn
ga
//
```

A.31 *HAECO*_S09_SUPERCONTIG_0037674 ≈ *CAEEL*-PGP-10

```
genewise
Query protein:       CAEEL-PGP-10___CAEEL-PGP-10 C54D1 1 1-1382AA
Target Sequence      H_S09_0037674___HAECO_S09_Supercontig_0037674

H_S09_0037674__-2071          Intron     TAG
                     <0-----[     : 1403]-0>


CAEEL-PGP-10___   793 LPGKETNFTVLKDLIYSYRTGLPLLAGAIPTTIVRAVFYLLICFQVASV
                      LPGK TNF+ +  +I +YR G  LLAGAIPTTI+RA FYLLICF+VASV
                      LPGKSTNFSAVWKVIANYREGYALLAGAIPTTILRAFFYLLICFEVASV
H_S09_0037674__-1402  ccgataataggtagagatcggtgtcgggacaaatcgttttcattgggtg
                      tcgaccatgctgattcaagagacttcgctcccttgcttatttgtatcct
                      gaaacccctatgatttcttattcgcgtatgcctgatcccaatttaccca


CAEEL-PGP-10___   842 LEI                                              844
                      LE+
                      LEV
H_S09_0037674__-1255  cgga                                            1246
                      tat
                      tag
```

```
H_S09_0037674__-1245  GTTGACC                                          1
                      <1-----[1245 :     ]-1>
                         1 n [50   :   50]
//

Gene 1
EXONS 1402 1246
  Exon 1402 1246 phase 0
//

>H_S09_0037674___HAECO_S09_Supercontig_0037674
_1402-1246bp_AA
_CAEEL-PGP-10_793-844AA

LPGKSTNFSAVWKVIANYREGYALLAGAIPTTILRAFFYLLICFEVASVLEV
//

>H_S09_0037674___HAECO_S09_Supercontig_0037674

ctgccaggaaaatccaccaacttcagtgcagtttggaaagttattgctaactatcgtgaa
ggttatgccttgctcgcgggtgcaattccgaccaccattttgcgagctttcttctactta
ctaatttgttttgaagtcgcctccgtacttgaagtga
//
```

A.32 *HAECO*_S09_SUPERCONTIG_0037517 ≈ *CAEEL*-PGP-10

```
genewise
Query protein:       CAEEL-PGP-10___CAEEL-PGP-10_C54D1_1_1-1382AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0037517___HAECO_S09_Supercontig_0037517
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623

H_S09_0037517__    1          Intron     CAG
                    <2-----[     : 2126]-2>
                      10 n [1957 : 1966]

CAEEL-PGP-10___  839  ASVLEISIAPDEERALQIFIVAAIYTALIIVKTIFEALG
                      S+L+I+IAP EER  QIFIVAA+YTALII+KTIFEALG
                      GSLLKIAIAPAEERNDQIFIVAAVYTALIIIKTIFEALG
H_S09_0037517__ 2127  gtttaagagcgggcagcataggggtagcaaaaatggcg
                      gcttatctcccaagaaattttcctacctttacttactg
                    tctaagcgatacaacttgtcttatccaattccatacaatt


CAEEL-PGP-10___  878                        RIFIALYGHGFCKFMRNEMFRK
                                            R+FIALYGHGFC  MR+ MFRK
                                            RLFIALYGHGFCSCMRSIMFRK
H_S09_0037517__ 2245                        actagttgcgttttaaaaataa
                    GTGAGTT  Intron 1   CAGgtttctagagtgcgtggtttga
                    <0-----[2245 : 2303]-0>atccagtcttttgtgaccgcag


CAEEL-PGP-10___  900                        VLRHGAAYFDEERNSPGRLVHKVINE
                                            ++RHG AYFDEERNSPGR++ ++I +
                                            IMRHGCAYFDEERNSPGRILQRIITD
H_S09_0037517__ 2370                        aaccgtgttgggcaacgcatcaaaag
                    GTTTCTT  Intron 2   CAGttgaggcataaagagcggttagttca
                    <0-----[2370 : 2439]-0>tgccatatctaggtccgttaaactcc


CAEEL-PGP-10___  926 SSTLNEIMEQKLDMLIPGVVCS              L          F
                     SSTLN+IME KLD+LIP V+C                L          F
                     SSTLNKIMESKLDILIPAVICP              L          F
H_S09_0037517__ 2518 ttacaaaagtacgacacggatcc           [ctt]         t
                                            ccctaattacatatttccttgc GTAAGTC  Intron 3   CAGtt
                                            tgggtatgagagtttagaaatc <1-----[2585 : 2801]-1>tt


CAEEL-PGP-10___  950 SIVCALWINWKMALLCSFQFPAYFVIRILQIREGTK             985
                     S+  A++INWKMALLCSFQFPAYFVIRI+Q++EGTK
                     SLAAAMYINWKMALLCSFQFPAYFVIRIVQMKEGTK
H_S09_0037517__ 2807 tcgggataataagcttttctcgttgacagcaaggaaa          2916
                     ctccctatagatcttgctatccatttgttataagcag
                     gtttcgcctgggacgcgcatttctctgcgagaaaaa


H_S09_0037517__      GTGCGAT  Intron                                 3929
                      <2-----[2917 :      ]-2>
//

Gene 1
EXONS 2128 2916
  Exon 2128 2244 phase 0
  Exon 2304 2369 phase 0
  Exon 2440 2584 phase 0
  Exon 2802 2916 phase 1
//

Making a L in phase 1 intron


>H_S09_0037517___HAECO_S09_Supercontig_0037517
_2127-2916bp_AA
_CAEEL-PGP-10_839-985AA

G
SLLKIAIAPAEERNDQIFIVAAVYTALIIIKTIFEALGRLFIALYGHGFCSCMRSIMFRK
IMRHGCAYFDEERNSPGRILQRIITDSSTLNKIMESKLDILIPAVICPLFSLAAAMYINW
KMALLCSFQFPAYFVIRIVQMKEGTK
//

>H_S09_0037517___HAECO_S09_Supercontig_0037517
_2127-2916bp

tggc
tctttattaaagatcgcgatagctccagccgaagaacgcaatgatcagattttcattgtt
gcagctgtctacacagcacttattatcatcaaaactatattcgaagcacttggtagactt
ttcatcgcattgtatggccatggtttttgttcgtgtatgagaagcatcatgttcagaaag
attatgcgccacggatgtgcatatttcgatgaagagcggaatagccccgggcgtatttta
caaagaatcattaccgactcttcgacgctgaataaaattatggaatcgaaactggatatt
cttataccggcagtaatatgtccccttttttcgcttgctgctgccatgtacatcaattgg
aagatggcactcttgtgctcgttccaatttcctgcttactttgtcattcggatcgtgcaa
atgaaagaaggaacaaaaag
//
```

A.33 *HAECO*_S09_SUPERCONTIG_0024002 ≈ *CAEEL*-PGP-11

```
genewise
Query protein:      CAEEL-PGP-11___CAEEL-PGP-11_DH11_3_1-1270AA
Target Sequence     H_S09_0024002___HAECO_S09_Supercontig_0024002

H_S09_0024002__    1             Intron CAG
                                 <0-----[     : 5288]-0>


CAEEL-PGP-11___   64 INFAKKSDLIFIITALICALLGGTIQPVVLLIGGWITDLYLTNGN
                       FA   D+ FI    +CAL+GG IQP VL++GG+IT +YL
                      FRFATTRDMCFITIGAVCALIGGAIQPFVLILGGFITTVYLEPSE
H_S09_0024002__ 5289 tctgaacgattaaagggtgcagggacctgcatggtaaagttgcag
                      tgtcccgatgttctgctgctttggctacttttttggttcctatacga
                      tatctaatgttagcgcgccctatttaactgcataccaagcggctg


CAEEL-PGP-11___  109                          TAGNDEFLYSVLTLI-YAGLGFGVII
                                               N +F   V+ LI   G+  GV+
                                             KVANQQFWDDVMVLINWLGIA-GVLA
H_S09_0024002__ 5424                         aggaccttgggaqcaatcgag ggcg
                      GTAGGCT  Intron 2   TAGatcaaatgaatttttagtgtc gttc
                      <0-----[5424 : 5485]-0>agacggtgtttgtgatgcata caat


CAEEL-PGP-11___  134 LVLALIQ                          GVCIQRGTSRILDSIRKEF
                      L+ + +Q                            + RG+  +++S+R+++
                      LITSFVQ                          SFFLHRGSLNVVNSLRRQY
H_S09_0024002__ 5561 caattgc                          tttcccgacaggatcccct
                      ttccttaGTAATAG  Intron 3   TAGctttagggtattactggaa
                      gaagttg<0-----[5582 : 6449]-0>cctctttcgcgctgtacgt
                             10 n [5711 : 5720]
                              1 n [5812 : 5812]


CAEEL-PGP-11___  160 LGAVLRQDANWLDKHSSGSITCQLNENI              E
                      L AVLRQDA W D+++SG IT QLNE +
                      LSAVLRQDATWFDQNTSGVITSQLNEFV             A
H_S09_0024002__ 6507 ttggtccggattgcaatggaatctagtgg           [gca]
                      tccttgaaccgtaaaccgttccataatt GTTCCTA  Intron 4
                      gtttatgtaagttgctgtccagggtgca <1-----[6592 : 7261]


CAEEL-PGP-11___  188     VISDGLGNKCCMLVRGFAMFTSSIIACAFINWQLTFITFTM
                         + S  L     +L +    +F S  I+    +     T    T+
                         LKSQCLIINTLLLFKAPYQFLSVAIS-SYHRVPTTTSHPTL
H_S09_0024002__ 7259     caacttaaaactctagctctttggaa ttcagcaaatccat
                      CAGctagagtttacttttaccaattctctg caagtcccccacct
                      -1>aagcgtgcatctgctgcttacaattct ctcaatgcgattag


CAEEL-PGP-11___  230                G          PVSAF                 235
                                                  +
                                    P          TSKIM
H_S09_0024002__ 7384 c           [ccg]         ataaa                 8403
                      cGTCATTT  Intron 5   GAG ccatt
                       <2-----[7386 : 8387]-2>gtcacg


CAEEL-PGP-11___  236 VLHL                                            239

                      Gxxx
H_S09_0024002__ 8404 gNNN                                           8415
                      gNNN
                      NNNN
                             10 n [8406 : 8415]


CAEEL-PGP-11___  240 LTKVNEVSNEELMSLSAQSHAIIEESILNVRTVQSCNGQNFMIT
                       +VN  S  + M L   + +I+EES++NV+TVQSCNGQ  M+
                      KFQVNSASLGKAMRLLTDAASIVEESVMNVKTVQSCNGQKHMVK
H_S09_0024002__ 8416 atcgatgatgagaatcagggtagggtgaagaagcttagcacaga
                      atataccgtgactgttcacccttaacttatactacgagaaatta
                      atggtttcacagggagattttataaatgtgatgggtcagatggg


CAEEL-PGP-11___  284                          KLNQVNEKIKKFYNKSTFWAGFFDGLA
                                             K      +   +  ++ FW GFF+GL+
                                             KYQAALSRALPYSIRNHFWMGFFEGLS
H_S09_0024002__ 8548                         atcggcacgccttacacttagttggct
                      GTCAGTG  Intron 6   CAGaaacctggctcactgaatgtgttagtc
                      <0-----[8548 : 9292]-0>atgatatttggtctcttcggttcggaa
```

```
CAEEL-PGP-11___  311 LFVIYFITGISL              F          FGCRLYFNQEIG
                       F IY + G+SL              +          FGC  YF+  +
                      FFQIYVVIGLSL              W          FGCYGYFHGFVR
H_S09_0024002__ 9374 ttcatggagctct           [tgg]        tgttgttcgtgc
                      ttatatttgtctgGTGACTT  Intron 7   CAG tggagatagttg
                      ctgccgctatct <2-----[9412 : 9514]-2>gcatcctcttttt


CAEEL-PGP-11___  336 KAGDVILIVNTICVTGYFLGLLGPHMSSLQQAATSPLLLYKTIES  380
                      + GDV+L V TI +T Y+LG+LGPHM +L +A  +   +Y+ I+
                      ERGDVLLCVGTISLTAYYLGMLGPHMMALLKARVAAAAIYEIIDR
H_S09_0024002__ 9552 gcgggcctggaattagttcgacgccaagctagaggggatgaaga 9686
                      aggatttgtgctctccaatgttgcattcttacgtcccctaattag
                      acttgtctgcatgactttgagtatcggtgggagagttcctaatcg


H_S09_0024002__ 9687 GTATGAG  Intron                                10697
                      <0-----[9687 :     ]-0>
                        10 n [9791 : 9800]
//

Gene 1
EXONS 5289 9686
  Exon 5289 5423 phase 0
  Exon 5486 5581 phase 0
  Exon 6450 6591 phase 0
  Exon 7262 7385 phase 1
  Exon 8388 8403
       8404 8415
       8416 8547 phase 2
  Exon 9293 9411 phase 0
  Exon 9515 9686 phase 2
//

Making a A in phase 1 intron
Making a P in phase 2 intron
Making a W in phase 2 intron


>H_S09_0024002___HAECO_S09_Supercontig_0024002
_5289-8403bp_AA 8404-8415bp_AA 8416-9686bp_AA
_CAEEL-PGP-11_64-235AA-236-239AA_240-380AA

FRFATTRDMCFITIGAVCALIGGAIQPFVLILGGFITTVYLEPSEKVANQQF
WDDVMVLINWLGIAGVLALITSFVQSFFLHRGSLNVVNSLRRQYLSAVLRQDATWFDQNT
SGVITSQLNEFVALKSQCLIINTLLLFKAPYQFLSVAISSYHRVPTTTSHPTLPTSKIM

G

KFQVNSASLGKAMRLLTDAASIVEESVMNVKTVQSCNGQKHMVKKYQAALSRALPYS
IRNHFWMGFFEGLSFFQIYVVIGLSLWFGCYGYFHGFVRERGDVLLCVGTISLTAYYLGM
LGPHMMALLKARVAAAAIYEIID
R
//

>H_S09_0024002___HAECO_S09_Supercontig_0024002

tttcgatttgccactacacgagatatgtgttttataacg
atcggggccgtgtgcgccctcattggaggtgctattcaaccattcgttctgatcttaggt
ggattcatcacaacagtgtacttggagcccagtgagaaagtggcaaaccagcagttttgg
gatgatgttatggttctgataaattggctcggaattgcaggcgtactagctctgataaca
tcgtttgttcagtcctttctttctccatcgtggtagcctgaacgtggtcaattcgcttcga
cgccagtatttgtctgctgttttacgtcaggatgcaacatggtttgatcagaacacttcg
ggtgtcatcacatcgcagttgaatgagttcgtagcactaaagagccagtgtttgatcata
aatacccttttgctctttaaggcccctatcaattcttatcagttgctatcagttcctat
cacagagtacctacgaccacgtcacatcctacattgccgacttccaaaatcatg

gg

aaatttcaggtgaattctgctagcttaggcaaagcgatgaggttactgacagat
gctgcttctatagttgaagaatcagttatgaatgtgaaaactgtgcagtcgtgtaacgga
cagaaacatatggtgaagaaatatcaggcagctctaagtcgtgctctgccgtattccatt
cgcaatcatttctggatgggttttttcgaggggctatcattctttcagatctacgtggtc
attggactttccctttggttcggatgttacggctatttccatggttttgttcgtgaacgc
ggtgatgtgcttctctgtgtgggcacaatttcgttaaccgcttattatctgggaatgctt
ggacctcacatgatggctctgttgaaggcaagggtagcggctgctgccatctatgaaata
attgac
agg
//
```

A.34 *HAECO*_S09_SUPERCONTIG_0024351 ≈ *CAEEL*-PGP-11

```
genewise
Query protein:      CAEEL-PGP-11___CAEEL-PGP-11 DH11 3 1-1270AA
Target Sequence     H_S09_0024351___HAECO_S09_Supercontig_0024351

H_S09_0024351_-11350                    CAT
                                  <?-----[     :11124]-?>


CAEEL-PGP-11___   301 FWAGFFDGLALFVIYFITGISL              F          F
                      FW GFF+GL+ F IY + G+SL                        F
                      FWMGFFEGLSFFQIYVVIGVSL             W          F
H_S09_0024351_-11123  ttagttggctttcatggaggtct           [tgg]        t
                      tgtgttagtcttatatttgtctgGTGATTT  Intron 1   TAG t
                      cggaccaggactgccgctattt <2-----[11055:10962]-2>gc


CAEEL-PGP-11___   325 GCRLYFNQEIGKAGDVILIVNTICVTGYFLGLLGPHMSSLQQAATSFLL
                      GC  YF+  + + GDV+L V TI +T Y+LG+LGPHM +L +A  +
                      GCYGYFHGFVRERGDVLLCVGTISLTAYYLGMLGPHMMALLKARVAAAA
H_S09_0024351_-10957  gttgttcgtgcgcgggcctggaattagttcgacgccaaqctagaqgggg
                      ggagatagttgaggatttgtgctctccaatgttgcattcttacgtcccc
                      gttttcttttacttgtctgcatcactttgagtatcggtgggagagttc


CAEEL-PGP-11___   374 LYKTIES                          APKSEGKEGEIKISSTRGN
                      +Y+ I+                            P    K+G ++     G+
                      IYEIIDR                          MPPIHRKKG-LRKDIVAGH
H_S09_0024351_-10810  atgaaga                         accaccaag caagagggc
                      taattagGTACGAG  Intron 2   CAGtcctagaag tgaattcga
                      ctaatcg<0-----[10789:10731]-0>gttatggat cagctgaat


CAEEL-PGP-11___   400 IEFRDVRFKYFTRDNEVLQ                          GLSLQVL
                      I ++DV F Y TR+  VL                           GLS +
                      IALKDVYFSYSTRNQPVLN                          GLSCEAQ
H_S09_0024351_-10676  agcaggtttttaaaccgca                          gcttggc
                      tctaatatcaccgaacttaGTGAGAA  Intron 3   CAGgtcgaca
                      tccatcccataaacaacgt<0-----[10619:10353]-0>ccgtata


CAEEL-PGP-11___   426 PGQTVALVGTSGCGKSTS                          IGLLTKLY
                       G+T+ALVG SG GKST                          + LL++LY
                      SGETIALVGASGSGKSTL                          VSLLSRLY
H_S09_0024351_-10331  aggaagcgggagtgaaat                          gaccacct
                      ggactcttgcggcgagctGTAATGA  Intron 4   CAGtgttggta
                      taagcgtctatattacag<0-----[10277:10200]-0>tctctgct


CAEEL-PGP-11___   452 RASEGEILIDGKNIDMLDAKSLRQ                          QI
                         +G+I IDG  +   + K+LR                          +I
                      EVDDGKITIDGVPLCDYEVKALRK                          RI
H_S09_0024351_-10175  ggggqaaaagggcctgtggagcaa                          aa
                      ataagatctagtctgaaatactgaGTGCCAA  Intron 5   CAGgt
                      agcctacctctgtctccaagggag<0-----[10103:10047]-0>ac


CAEEL-PGP-11___   482 GIVQQEPKLFDGTIMENIKLGR-NVDEETIKTAADIANASSFIEKLEN
                      GIVQQEP LF+GT+ ENI LGR  +D+E I+ AAD+A+ + F++ LE
                      GIVQQEPYLFNGTVKENISLGREGIDDEKIRQAADVADVTEFVDHLEK
H_S09_0024351_-10028  gagccgctctagagagaattgcggagggaaacgggggggagtggccga
                      gttaaacattagctaaatctggagtaaaatgaccatcatcattaataa
                      aatggagtacctaaggcaggagattctaacaatttgcctcattctcag


CAEEL-PGP-11___   525                          GYETRLGPGGVQLSGGQKQRICIARA
                                               G++T LGPGG  LSGGQKQRI IARA
                                               GFDTYLGPGGATLSGGQKQRIAIARA
H_S09_0024351_-9896                            gtgattgcggactggcaccagagcg
                      GTAATAC  Intron 6   CAGgtacatgcgggcctcggaaagtctcgc
                      <0-----[9896 : 9792]-0>tcctcgttcccttcatggagccctac


CAEEL-PGP-11___   551 LVTSPSILLLDEATSALDSHNEHIVN
                      +VT P IL LDEATSALD+  E +V
                      IVTDPRILFLDEATSALDAKCEKVVQ
H_S09_0024351_-9713   agagccacttgggaagcggatgaggc
                      ttcacgttttaaccgctacagaattaGTAATGT  Intron 7   CAG
                      tctttgctcgtatctcatcgtgattg<0-----[9635 : 8753]-0>
                                                  10 n [8837 : 8828]


CAEEL-PGP-11___   577 KALTKASEGRTTIIIAHRLSSLKSVDRIYVLDQGKTKEI
                       AL +A+EGRTTI+IAHRLS+++ V +IYV+++GK  E
                      VALNEAAEGRTTIVIAHRLSTIRDVKKIYVMEKGKVVES
H_S09_0024351_-8752   ggcaggggaaaagagcctaaacggaaatgagagagggt
                      tctaaccaggcctttcagtgctgataatattaagattac
                      gtgtaggaaaacccaattgtttttagaatcgagaaatga


CAEEL-PGP-11___   616              G          THDELILLGGIY          628
                                   G          +HDEL+  GG+Y
                                   G          SHDELMNKGGLY
H_S09_0024351_-8635   g          [gga]        tcggcaaaggtt         7938
                       GTAAGAG  Intron 8   CAGgcaaattaaggta
                       <1-----[8634 : 7976]-1>aactaggtaacgt


CAEEL-PGP-11___   629 ARLAKSQEVE                                      638
                      A++   Q+ +
                      AQMVTVQQFQ
H_S09_0024351_-7937   gcagagcctca                                     7909
                      cattctaata
                      ggggatgata


H_S09_0024351_-7908 GTAGACC  Intron                                     1
                    <1-----[7908 :     ]-1>
//

CAEEL-PGP-11     639 QSSKKDWEREELRRAEKMKKKGRTVEIIEPNSTILQEHEHNFVGSVITEN
H_S09_0024351_-     ----------------------------------------------------


CAEEL-PGP-11___   689 EEQKISFSGISKLFNYLPKHRKTLILIIAALLIPR              722
H_S09_0024351_-      ----------------------------------


H_S09_0024351_-           Intron     CAG
                      <1-----[     : 6276]-1>


CAEEL-PGP-11___   723  AIELCSYGIGMSFAFKTLQ                           741
                       A+EL   G+   FAFK+LQ
                       ALELPGLGLAYLFAFKSLQ
H_S09_0024351_-6273    gcgccgcgcgtctgtattc                          6216
                       ctatcgtgtcattctacta
                      agaatttaatcttcccgtag


H_S09_0024351_-6215 GTACGAG  Intron
                    <0-----[6215 :      ]-0>
//

CAEEL-PGP-11     742 RSKDDYMTWNYITLAQQTLAGITFWILHTSL                 772
H_S09_0024351_-     --------------------------------

Alternative splice

H_S09_0024351_-           Intron     CAG
                      <0-----[     : 5428]-0>


CAEEL-PGP-11___   773 MYLCGWLANEVMNEVKQEMLSEVLNKPIPYFDNPETSPSACVSRIISHAH
                       L GWL+ +VM+ ++  +L  +L++P+ YFD  ETSP++CV+ +  HA
```

```
                      FSLSGWLSEQVMDGLRARILKSLLHRPMTYFDCEETSPASCVATVSQHAP
H_S09_0024351_-5427   ttctgtcagcgaggtagcacaacccacaattgtggatcgttggagtccgc
                      tctcggtgaattagtgcgttagttagctcatagaaccccgtcctcaacc
                      ctttcgatgaagtaggagaggtgctagggtcttaaggaattgcaaagttc


CAEEL-PGP-11___   823 NCYA                          CLDHRAIRFVWFIAGTIFSLLLA
                      +  A                           LD+R +  +  +A +++ +++A
                      HAMA                          ALDYRLMTNLSNLAASVIGIIIA
H_S09_0024351_-5277   cgag                          gcgtctaaacaatggagagaaag
                      actcGTAAGTC  Intron 1   TAGctaagttcatgatccgttgtttc
                      tcgg<0-----[5265 : 5093]-0>acctaagctgccggttcctctat


CAEEL-PGP-11___   850 FPFVWELGVLGL                                    871
                        F W LG+LG
                      VTFSWWLGLLGA
H_S09_0024351_-5023   gatttcgccggg                                   4987
                      tctcggtgtgc
                      cgttggaactaa


H_S09_0024351_-4986 GTAATTT  Intron
                    <1-----[4986 :     ]-1>
                     660 n [4360 : 3701]
//

Alternative splice

H_S09_0024351_-           Intron     CAG
                      <0-----[     : 3369]-0>


CAEEL-PGP-11___   773 MYLCGWLANEVMNEVKQEMLSEVLNKPIPYFDNPETSPSACVSRIISHAH
                       L GWL+ +VM  ++  +L  +L++P+ Y+D  ETSP++CV+ +  HA
                      FSLSGWLSEQVMAGLRARILKSLLHRPMTYYDCEETSPASCVATVSQHAP
H_S09_0024351_-3368   ttctgtcagcgaggtagcacaacccacaattgtggatcgttggagtccgc
                      tctcggtgaattcgtgcgttagttagctcaaagaaccccgtcctcaacc
                      ctttcgatgaagtaggagaggttctagggtcttaaggaatcgcaaggttc


CAEEL-PGP-11___   823 NCYA                          CLDHRAIRFVWFIAGTIFSLLLA
                      +  A                           LD+R +  +  +A +++ +++A
                      HAMA                          ALDYRLMTNLSNLAASVIGIIIA
H_S09_0024351_-3218   cgag                          gcgtctaaacaatggagagaaag
                      actcGTAAGTC  Intron 1   TAGctaagttcatgatccgttgtttc
                      tcgg<0-----[3206 : 3038]-0>acctaagctgccggttcctctat


CAEEL-PGP-11___   850 FPFVWELGVLGL                                    871
                        + W LG+LG
                      VTYSWWLGLLGA
H_S09_0024351_-2968   gattttcgccggg                                  2932
                      tcacggtgttgc
                      cgttggaactaa


H_S09_0024351_-2931 GTAATTT  Intron
                    <1-----[2931 :     ]-1>
                     10 n [1908 : 1899]
//

Gene 1
EXONS 11123 2930
  Exon 11123 11056 phase 0
  Exon 10961 10790 phase 2
  Exon 10730 10620 phase 0
  Exon 10352 10278 phase 0
  Exon 10199 10104 phase 0
  Exon 10046 9897 phase 0
  Exon 9791 9636 phase 0
  Exon 8752 8635 phase 0
  Exon 7975 7909 phase 1
//

  ---

  Exon 6273 6216 phase 0
//

  ---
  Exon 5427 5266 phase 0
  Exon 5092 4987 phase 0
//

  Exon 3368 3207 phase 0
  Exon 3037 2932 phase 0
//

Making a W in phase 2 intron
Making a G in phase 1 intron


>H_S09_0024351___HAECO_S09_Supercontig_0024351
_11123-7909bp_AA
---
_6273-6216bp_AA
---
_5427-4987bp_AA_Alternative_splice
_3368-2932bp_AA_Alternative_splice
_CAEEL-PGP-11_301-638AA---723-741AA---773-871AA_Alternative_splices

FWMGFFEGLSFFQIYVVIGVSLWFGCYGYFHGFVRERGDVLLCVGTISLTAYYLGMLGPH
MMALLKARVAAAAIYEIIDRMPPIHRKKGLRKDIVAGHIALKDVYFSYSTRNQPVLNGLS
CEAQSGETIALVGASGSGKSTLVSLLSRLYEVDDGKITIDGVPLCDYEVKALRK
RIGIVQQEPYLFNGTVKENISLGREGIDDEKIRQAADVADVTEFVDH
LEKGFDTYLGPGGATLSGGQKQRIAIARAIVTDPRILFLDEATSALDAKCEKVVQVALNE
AAEGRTTIVIAHRLSTIRDVKKIYVMEKGKVVESGSHDELMNKGGLYAQMVTVQQFQ
//

ALELPGLGLAYLFAFKSLQ
//

FS
LSGWLSEQVMDGLRARILKSLLHRPMTYFDCEETSPASCVATVSQHAPHAMAALDYRLMT
NLSNLAASVIGIIIAVTFSWWLGLLGA
//

FS
LSGWLSEQVMAGLRARILKSLLHRPMTYYDCEETSPASCVATVSQHAPHAMAALDYRLMT
NLSNLAASVIGIIIAVTYSWWLGLLGA
//
_11123-7909bp

ttctggatgggattcttcgaagggctgtcattctttcagatctacgtggtcattggagtt
tctctttggttcgggtgttatggttatttccatggttttgttcgtgaacgcggtgatgtg
cttctctgtgtgggcacaatttccttaaccgcttattatctgggaatgcttggacctcac
atgatggctctgttgaaggcaagggtagcggctgctgccatctatgaaataattgacagg
atgcctcctatacatcggaagaaaggtctcagaaaggacattgtggcaggacatattgcc
ctcaaagatgtctacttctcatattcaacaagaaaccaaccagtcctgaatggcctctcg
tgtgaagctcaaagtggagaaacgatcgcgcttgtcggtgcaagtggatctggtaaaagc
acattggttagccttctcagtcggctctatgaagtggacgacggtaaaatcaccattgac
ggtgtgcctctctgtgactacgaagtaaaggcgctgagaaag
agaatcggaatagttcagcag
gaaccgtatctattcaacggtacagtaaaggagaacatatcgttgggacgggaaggtatt
gacgatgaaaaaatcagacaagctgctgatgtggccgacgttaccgaatttgttgaccat
ctcgaaaaggtttcgacacttacttgggtcctggcggcgccactctttccggaggtcag
aagcaacggatcgccatcgctcgagccattgtcactgatcctcggatcctttttcttggat
gaagctaccagtgccctagatgccaagtgtgagaaagttgttcaggtggctctgaatgaa
```

gcggcggaaggaagaacaaccatcgtcatagcacatcgtttgagtactattcgtgatgta
aagaaaatatatgtcatggaaaagggaaaagtagttgagtcaggatcacacgatgaactg
atgaataaaggaggcttgtatgcgcagatggtgacagttcagcaatttcaaa
//

>H_S09_0024351___HAECO_S09_Supercontig_0024351
_6277-6220bp

agcgctagaacttcctggtctaggacttgcctatcttttcgccttcaagtctttacag
//

>H_S09_0024351___HAECO_S09_Supercontig_0024351
_5427-4987bp

ttctct
ctttctggctggctaagtgagcaagtaatggatggattgagggcacggatactgaagagt
ctgctccatagaccgatgacgtatttcgattgtgaagaaacgtcgccagcatcttgtgtg
gccacagtatcacagcatgctccccatgccatggcggcactcgactatcgattaatgacc
aatctgagcaacttggcggctagtgtcatcggtatcattatagctgtcacgttttcttgg
tggctaggactccttggagcag
//

>H_S09_0024351___HAECO_S09_Supercontig_0024351.[3362:2928]
_3368-2932bp

ttctct
ctttctggctggctaagtgagcaagtaatggctggattgagggcacggatactgaagagt
cttctccatagaccgatgacgtattacgattgtgaagaaacgtcgccagcatcttgcgtg
gccacagtatcgcagcatgctccccatgccatggcggcactcgactatcgattaatgacc
aatctgagcaacttggcggctagtgtcatcggtatcattatagctgtcacgtattcttgg
tggctaggactccttggagcag
//

A.35 *HAECO*_S09_SUPERCONTIG_0035404 ≈ *CAEEL*-PGP-11

genewise

```
H_S09_0035404__ 1009          Intron     TAG
                <0-----[     :  906]-0>


CAEEL-PGP-11___ 1047 VCEGFSLNIPK
                     V +  +L
                     VAQNLNLMARC
H_S09_0035404__  905 ggcacacagct
                     tcaatattcgg
                     tcatatcgaac


CAEEL-PGP-11___ 1058 GHSIALVGASGCGKSTIISMLERFYSAKAGRI
                     G +IALVGASGCGKST+I +LERFY   +G I
                     GQAIALVGASGCGKSTVIQLLERFYEPDSGNI
H_S09_0035404__ -872 gcgagcgggtgtgaaagacccgattgcgagaa
                     gactcttgccgggagcttattagtaacaggat
                     tgcactatagctggtaacgtggaccagctttt


CAEEL-PGP-11___ 1090                       SVDDNDIDGIDVNHLRNNISVVGQEP
                                           +D++++  +   HLRNNI++VGQEP
                                           KIDNHELKQLCRVHLRNNIALVGQEP
H_S09_0035404__ -776                       aagacgcacttcgcccaaagcggcgc
                     GTGAGTT  Intron 1   CAGataaaataatggtatgaatcttgaac
                     <0-----[776  :  686]-0>accccgggggtctctgcccccagggt


CAEEL-PGP-11___ 1116 VLFN              A          TIRENITIGIDEVSVEEVQ
                     +LF                         +I ENIT+G+++VSV EVQ
                     ILFK              G         SIIENITLGLEDVSVAEVQ
H_S09_0035404__ -607 actag          [ggg]        taagaaacgcgggaggggc
                     ttta GTGATAC  Intron 2   TAGgcttaatctgtaatgtcata
                     ttca <1-----[594  :  543]-1>ggccactggagatctacgag


CAEEL-PGP-11___ 1140 KACKAANAAGFIESFPL                        GYDTIVGEG
                     +AC+ ANAA F+E+FP                         GY+T VGE
                     EACRQANAANFVEAFPQ                        GYETDVGEK
H_S09_0035404__ -483 ggtacgaggatgggtcc                        gtgagggga
                     acggacaccattactcaGTCCTGA  Intron 3   CAGgaacatgaa
                     ggcgggtgccttactcg<0-----[432  :  374]-0>acatttcgg


CAEEL-PGP-11___ 1166 GASLSGGQKQRIAIARAIIRKPKILLLDEATSALDTQSEE
                     G SLSGGQKQRIAIARA+IRKPK++LLDEATSALDT+SE+
                     GGSLSGGQKQRIAIARALIRKPKVILLDEATSALDTESEK
H_S09_0035404__ -346 ggactggcaccagagagtacacagaccgggaagcgagaga
                     gggtcggaaagtctcgcttgacattttaaccgctacagaa
                     accaattggatccacatgtcatacagatatttatccacag


CAEEL-PGP-11___ 1206                        IVQKALRSATTGRTSIIVAHRLSTVQ
```

```
                                           +VQ AL  A+ GRTSI +AHRLSTV+
                                           VVQNALNEASHGRTSITIAHRLSTVR
H_S09_0035404__ -226                       ggcagtaggacgaataaagcataaga
                     GTCACTA  Intron 4   AAGttaactaacgaggcctctcagtgctg
                     <0-----[226  :   94]-0>gtatgacaactcaggtaaatagtcta


CAEEL-PGP-11___ 1232 HCD                                            1234
                       D
                     DAD
H_S09_0035404__  -15 ggga                                             5
                     acag
                     ctt


H_S09_0035404__    4 GTAA     Intron                                  1
                       <2-----[4    :     ]-2>
//

Gene 1
EXONS 905 5
  Exon 905 777 phase 0
  Exon 685 595 phase 0
  Exon 542 433 phase 1
  Exon 373 227 phase 0
  Exon 93 5 phase 0
//

Making a G in phase 1 intron


>H_S09_0035404___HAECO_S09_Supercontig_0035404
_905-5bp_AA

VAQNLNLMARC
GQAIALVGASGCGKSTVIQLLERFYEPDSGNIKIDNHELKQLCRVHLRNNIALVGQEPIL
FKGSIIENITLGLEDVSVAEVQEACRQANAANFVEAFPQGYETDVGEKGGSLSGGQKQRI
AIARALIRKPKVILLDEATSALDTESEKVVQNALNEASHGRTSITIAHRLSTVRDAD
//

>H_S09_0035404___HAECO_S09_Supercontig_0035404
_905-5bp

gttgcccaaaatctaaatctcatggcacgatgc
ggtcaggccatagcccttgtaggtgcatcgggctgtgggaagagtacagtaatccagctt
ctggagagattctacgaaccggacagtggtaatattaaaatcgacaaccacgagctgaag
cagttgtgtcgcgttcaccttcggaacaacatcgccctcgtagggcaggagcctattctt
ttcaaagggtcgatcatcgaaaacattacgctgggactggaagatgtcagtgtagccgag
gtacaggaggcgtgcaggcaggcgaatgcggccaactttgttgaagccttttccccaggga
tacgaaactgatgttggcgagaagggaggcagcctatcaggtggtcagaagcaacgtatc
gccatagccagagctttgattcgcaaacctaaagtcatactgctagatgaagctactagt
gcacttgacaccgaaagcgaaaaggtggttcaaaatgcgttaaacgaagcaagccatggc
agaacgtcgattacaatagcacatagattgagtaccgttagagacgctgat
ag
//
```

A.36 *HAECO*_S09_SUPERCONTIG_0069412 ≈ *CAEEL*-PGP-14

```
genewise
Query protein:      CAEEL-PGP-14___CAEEL-PGP-14 F22E10 3 1-1327AA
Target Sequence     H_S09_0069412___HAECO_S09_Supercontig_0069412

H_S09_0069412__    1          Intron     CAG
                     <2-----[     :  297]-2>


CAEEL-PGP-14___  127  GRVTNALLVYPPTSKQFRNKANENVYIFLGIGIFISITNFIQ
                      GRVTN LLVYPP SK+FRN+A ENVYIFLGIG+F+ ITNFIQ
                      GRVTNVLLVYPPNSKEFRNEAYENVYIFLGIGVFVLITNFIQ
H_S09_0069412__  298  gcgaagccgtccatagtcaggtgagtatcgaggtgcaaatac
                      ggtcattttaccacaatgaacaaatatttgtgtttttcatta
                     taatgttgcttaataaacttaccatattcctccccccaccag


CAEEL-PGP-14___  169                        YMCFQHCCTRVMAQMRHRFVYSVLRQ
                                           +MCF  CCTRV+++MRH++V ++LRQ
                                           FMCFHSCCTRVISKMRHEYVRAILRQ
H_S09_0069412__  425                       tattcattacgataaccgtgcgaccc
                     GTTATAT  Intron 1   CAGttgtagggcgttcatgaaatgcttga
                     <0-----[425  : 1043]-0>tgctccctttatgagtcattagactg


CAEEL-PGP-14___  195 NAGWFDKNHSGTITTKLN                              212
                     NAGWFD+NHSG ++TKLN
                     NAGWFDRNHSGALSTKLN
H_S09_0069412__ 1122 aggttgaactggctaatag                            1177
                     acggtagaacgctccataa
                     tacgtcgtccgggaaagc
```

```
H_S09_0069412__ 1178 GTGAGTC                                        1186
                       <2-----[1178  :     ]-2>

//

Gene 1
EXONS 298 1177
  Exon 298 424 phase 0
  Exon 1044 1177 phase 0
//

>H_S09_0069412___HAECO_S09_Supercontig_0069412
_298-1177bp_AA
_CAEEL-PGP-14_127-212AA

GRVTNVLLVYPPNSKEFRNEAYENVYIFLGIGVFVLITNFIQFMCFHSCCTRVISKMRH
EYVRAILRQNAGWFDRNHSGALSTKLN
//

>H_S09_0069412___HAECO_S09_Supercontig_0069412

tggacgagttacgaatgttctgctcgtttatccaccaaattcaaaagaattccgtaat
gaagcctacgaaaatgtatatattttcctcggtatcggcgtcttcgtcctcatcacaaac
ttcatacagtttatgtgctttcacagctgctgtactcgtgtaatttcgaaaatgcgtcac
gaatatgttcgagcgatactccgtcagaatgcaggctggtttgacaggaatcactccggg
gcgctgtcaacaaaattgaacga
//
```

A.37 *HAECO*_S09_SUPERCONTIG_0004549 ≈ *CAEEL*-PGP-14

```
genewise
Query protein:       CAEEL-PGP-14___CAEEL-PGP-14 F22E10 3 1-1327AA
Target Sequence      H_S09_0004549___HAECO_S09_Supercontig_0004549

H_S09_0004549__-3157          Intron    CAG
                       <2-----[     : 880]-2>
                         10 n [1685 : 1676]


CAEEL-PGP-14___  214  SMERIREGIGDKLGVLLRGFAMLIAAIVVAYIYEWRLASMMLGVAPTCC
                      +MERIREGIGDKLG+LLRG AM  AA+++A+IYEWRLA MMLGV PT C
                      NMERIREGIGDKLGLLLRGCAMFTAAVIIAFIYEWRLALMMLGVTPTTC
H_S09_0004549__ -879  aagaacggaggacgctcagtgatagggaagtatgtctgtaacggacaat
                      atagtgagtgaatgtttgggcttccctttcttaaggtcttttgtccccg
                      ccggattaacctattaggaatcgctatgttacctagagagggtgggacgt


CAEEL-PGP-14___  263 ICMSLLAR                                        270
                       MS++AR
                     AIMSIMAR
H_S09_0004549__ -731 gaataaga                                        708
                     cttcttcg
                     ctgctgca
```

```
H_S09_0004549__ -707 GTAAGTC  Intron                                  1
                     <0-----[707  :     ]-0>
//

Gene 1
EXONS 879 708
  Exon 879 708 phase 0
//

>H_S09_0004549___HAECO_S09_Supercontig_0004549
_879-708bp_AA
_CAEEL-PGP-14_214-270AA

NMERIREGIGDKLGLLLRGCAMFTAAVIIAFIYEWRLALMMLGVTPTTCAIMSIMAR
//

>H_S09_0004549___HAECO_S09_Supercontig_0004549

c
aacatggagagaattcgtgaaggaatcggcgataaacttggtctattgctgagaggatgt
gccatgttcactgcagctgtgattattgcattcatctatgaatggcgattggcattgatg
atgcttggggtgacgccaaccacgtgtgccattatgtccattatggccaga
//
```

A.38 *HAECO*_S09_SUPERCONTIG_0059902 ≈ *CAEEL*-PGP-13

```
genewise
Query protein:       CAEEL-PGP-13___CAEEL-PGP-13 F22E10 2 1-1324AA
Target Sequence      H_S09_0059902___HAECO_S09_Supercontig_0059902

H_S09_0059902__-1094          Intron     CAG
                               <2-----[     :  182]-2>


CAEEL-PGP-13___  340  YGGYLLKVGIIKSPGDVFIIVVAMLLGAYFLGLISPHLM      378
                      YGGYLLKV IIK+PGDVFI+V+++LLGAYFLGLISPHLM
                      YGGYLLKVNIIKTPGDVFIVVMSLLLGAYFLGLISPHLM
H_S09_0059902__ -181  tggttcagaaaaacgggtaggatcccggttcgcatccta      64
                      aggattatattaccgatttttttctttgcattgttccatt
                     atttgcagccaaatctgtttggtaatccttgagctgtag


CAEEL-PGP-13___  379 V                                           379
                     !!!!
H_S09_0059902__  -63 gtac                                       60


CAEEL-PGP-13___  380 LLNARVAAASIYKTIDR                          396
                     LLNARVAAA+IY+TIDR
                     LLNARVAAATIYQTIDR
H_S09_0059902__  -59 ccagaggggaatcaagc                           9
                     ttacgtcccctaactag
                     gctaagattcccgttcg


H_S09_0059902__   -8 GTAGGAA  Intron                            1
```

```
                                        <0-----[8     :     ]-0>
//

Gene 1
EXONS 181 9
  Exon 181 64
         63 60
         59  9 phase 0

//

>H_S09_0059902___HAECO_S09_Supercontig_0059902
_181-9bp_AA
_CAEEL-PGP-13_340-378AA_379AA_Frameshift_380-396AA

YGGYLLKVNIIKTPGDVFIVVMSLLLGAYFLGLISPHLMXLLNARVAAATIYQTIDR
//

>H_S09_0059902___HAECO_S09_Supercontig_0059902

a
tatggtggttatttgctcaaagtgaacatcataaaaacacctggcgatgtgtttattgtt
gtgatgtctctactacttggcgcctattttctgggactgatctctccgcatttaatg

>H_S09_0059902___HAECO_S09_Supercontig_0059902

gtac

ctgctcaatgcaagagtggcagctgctaccatctaccagactattgaccgg
//
```

A.39 *HAECO*_S09_SUPERCONTIG_0021658 ≈ *CAEEL*-PGP-12

```
genewise
Query protein:       CAEEL-PGP-12___CAEEL-PGP-12 F22E10 1 1-1318AA
Target Sequence      H_S09_0021658___HAECO_S09_Supercontig_0021658

H_S09_0021658__     1          Intron     CAG
                               <0-----[     :  374]-0>


CAEEL-PGP-12___   501 EPILFNDTIHNNLLIGNPSATREDMIRVCKMANAHDFIQKMPN
                      EP LFNDT+  NL +GNP+ + E M+ VCKMANAHDFI K+PN
                      EPCLFNDTVAGNLRMGNPTMSLEQMVYVCKMANAHDFIGKLPN
H_S09_0021658__   375 gctctagagggaccagacaattgcagtgtaagagcgtagacca
                      acgttaactcgatgtgacctctaattatgatcacaattgatca
                      gactcttagattttgctatgcgaaggcacagatactctcaggt


CAEEL-PGP-12___   544                          GYETMIGDGSVQLSGGQKQRVAIART
                                               YET IGDG VQLSGGQKQR+AIART
                                              AYETYIGDGGVQLSGGQKQRIAIART
H_S09_0021658__   504                          gtgatagggggcctggcaccagagca
                      GTAAGTT  Intron 1   CAGcaacatgaggtatcggaaagtctcgc
                      <0-----[504  : 1062]-0>ccgcctccgtgggactgaagcctata


CAEEL-PGP-12___   570 LIRDPK                         VLLLDEATSALDAQSE
                      L RDPK                         VLLLDEATSALDAQSE
                      LARDPK                         VLLLDEATSALDAQSE
H_S09_0021658__  1141 tgcgca                         gcccgggaagcggcag
                      tcgacaGTGGATT  Intron 2   TAGttttaaccgctacaga
                      gattag<0-----[1159 : 1213]-0>ttagcaaattcttata


CAEEL-PGP-12___   592 SVVQ                                          595
                      S+VQ
                      SIVQ
H_S09_0021658__  1262 aagc                                        1273
                      gtta
                      ctag


H_S09_0021658__  1274 GTATGTA  Intron                              1448
                      <0-----[1274 :     ]-0>
//

Gene 1
EXONS 375 1273
  Exon 375 503 phase 0
  Exon 1063 1158 phase 0
  Exon 1214 1273 phase 0
//

>H_S09_0021658___HAECO_S09_Supercontig_0021658
_375-1273bp_AA
_CAEEL-PGP-12_501-595AA

EPCLFNDTVAGNLRMGNPTMSLEQMVYVCKMANAHDFIGKLPNAYETYIGDGGVQLSGG
QKQRIAIARTLARDPKVLLLDEATSALDAQSESIVQ
//

>H_S09_0021658___HAECO_S09_Supercontig_0021658

gagccatgccttttcaatgatacagtggcaggtaatcttcgtatgggcaatccaact
atgtccttggaacaaatggtgtacgtatgcaaaatggcaaatgcacacgatttcattggc
aaactgccgaatgcctacgagacctacattggcgacgggggtgtgcagctgtcaggcggt
cagaaacaacggatcgccattgcacgtacattggcacgtgatccaaaggttcttctactg
gacgaagcaacaagtgctctcgatgctcaaagtgaaagcattgtacag
//
```

```
genewise
Query protein:       CAEEL-PGP-12___CAEEL-PGP-12 F22E10 1 1-1318AA
Target Sequence      H_S09_0021658___HAECO_S09_Supercontig_0021658

H_S09_0021658__     1          Intron     CAG
                               <0-----[     :  374]-0>


CAEEL-PGP-12___   501 EPILFNDTIHNNLLIGNPSATREDMIRVCKMANAHDFIQKMPN
                      EP LFNDT+  NL +GNP+ + E M+ VCKMANAHDFI K+PN
                      EPCLFNDTVAGNLRMGNPTMSLEQMVYVCKMANAHDFIGKLPN
H_S09_0021658__   375 gctctagagggaccagacaattgcagtgtaagagcgtagacca
                      acgttaactcgatgtgacctctaattatgatcacaattgatca
                      gactcttagattttgctatgcgaaggcacagatactctcaggt


CAEEL-PGP-12___   544                          GYETMIGDGSVQLSGGQKQRVAIART
                                               YET IGDG VQLSGGQKQR+AIART
                                              AYETYIGDGGVQLSGGQKQRIAIART
H_S09_0021658__   504                          gtgatagggggcctggcaccagagca
                      GTAAGTT  Intron 1   CAGcaacatgaggtatcggaaagtctcgc
                      <0-----[504  : 1062]-0>ccgcctccgtgggactgaagcctata


CAEEL-PGP-12___   570 LIRDPK                         VLLLDEATSALDAQSE
                      L RDPK                         VLLLDEATSALDAQSE
                      LARDPK                         VLLLDEATSALDAQSE
H_S09_0021658__  1141 tgcgca                         gcccgggaagcggcag
                      tcgacaGTGGATT  Intron 2   TAGttttaaccgctacaga
                      gattag<0-----[1159 : 1213]-0>ttagcaaattcttata


CAEEL-PGP-12___   592 SVVQ                                          595
                      S+VQ
                      SIVQ
H_S09_0021658__  1262 aagc                                        1273
                      gtta
                      ctag


H_S09_0021658__  1274 GTATGTA  Intron                              1448
                      <0-----[1274 :     ]-0>
//

Gene 1
EXONS 375 1273
  Exon 375 503 phase 0
  Exon 1063 1158 phase 0
  Exon 1214 1273 phase 0
//

>H_S09_0021658___HAECO_S09_Supercontig_0021658
_375-1273bp_AA
_CAEEL-PGP-12_501-595AA

EPCLFNDTVAGNLRMGNPTMSLEQMVYVCKMANAHDFIGKLPNAYETYIGDGGVQLSGG
QKQRIAIARTLARDPKVLLLDEATSALDAQSESIVQ
//

>H_S09_0021658___HAECO_S09_Supercontig_0021658

gagccatgccttttcaatgatacagtggcaggtaatcttcgtatgggcaatccaact
atgtccttggaacaaatggtgtacgtatgcaaaatggcaaatgcacacgatttcattggc
aaactgccgaatgcctacgagacctacattggcgacgggggtgtgcagctgtcaggcggt
cagaaacaacggatcgccattgcacgtacattggcacgtgatccaaaggttcttctactg
gacgaagcaacaagtgctctcgatgctcaaagtgaaagcattgtacag
//
```

A.40 *HAECO*_S09_SUPERCONTIG_0050285 ≈ *CAEEL*-PGP-12

```
genewise
Query protein:      CAEEL-PGP-12___CAEEL-PGP-12 F22E10 1 1-1318AA
Target Sequence     H_S09_0050285___HAECO_S09_Supercontig_0050285

H_S09_0050285__     1          Intron     CAG
                               <0-----[     :   59]0->


CAEEL-PGP-12___   593 VVQSALNNAAKGRTTIMIAHRLSTIREAD
                      VVQ AL NA +GRT I IAHRLS+++  D
                      VVQQALENAKRGRTVISIAHRLSSVQHYD
H_S09_0050285__    60 ggccgcgagacgcagaaagcctaagcctg
                      ttaactaacagggcttgtcagtggtaaaa
                      atgaaggcgaacttacccttcactgactt


CAEEL-PGP-12___   622                K          IVFFENGVIVESGNHEELVALGG
                                     +          I+++ENGV+ ESG H EL+ L G
                                     R          ILYIENGVVAESGTHSELIKLDG
H_S09_0050285__   147 c          [cgg]          actagaggggagactgcaacgg
                      gGTAGGTA  Intron 1   CAG ttataagttcaggcacattatag
                       <2-----[149  :  305]-2>gaactataaaagtatctaccagca


CAEEL-PGP-12___   646 RYAKLVEAQKFKESD                              660
                      +YA LVE+Q  K  D
```

```
                      KYAALVESQDTKAPD
H_S09_0050285__   376 atggcggacgaagcg                            420
                      aaccttagaacacca
                      atgactacatggact


H_S09_0050285__   421 [421  : 2931]                               2931
//

Gene 1
EXONS 60 420
  Exon 60 148 phase 0
  Exon 306 420 phase 2
//

Making a R in phase 2 intron


>H_S09_0050285___HAECO_S09_Supercontig_0050285
_60-420bp_AA
_CAEEL-PGP-12_593-660AA

VVQQALENAKRGRTVISIAHRLSSVQHYDRILYIENGVVAESGTHSELIKLDGKYAALVE
SQDTKAPD
//
```

A.41 *HAECO*_S09_SUPERCONTIG_0015622 ≈ *CAEEL*-PGP-14

```
genewise
Query protein:      CAEEL-PGP-14___CAEEL-PGP-14 F22E10 3 1-1327AA
Target Sequence     H_S09_0015622___HAECO_S09_Supercontig_0015622

H_S09_0015622__-2843          Intron     CAG                              2540
                              <0-----[     : 2540]-0>


CAEEL-PGP-14___   828 SVLFAVVSENLSMRFRVQSFRNLLYQ
                      SV FA+VSENL+MRFRV+SF+NLLYQ
                      SVFFAIVSENLAMRFRVESFKNLLYQ
H_S09_0015622__-2539 agttgagtgatgactcggttaacctc
                      gtttcttcaatctgtgtactaattaaGTAGGTA  Intron 1   CAG
                      tgccacgagtgggatagaccataccg<0-----[2461 : 2211]-0>


CAEEL-PGP-14___   854 DASYFDNPAHAPGKLITRLASDAPNIKA
                      DASYFDNPAH PGKLITRLASDAPNIKA
                      DASYFDNPAHTPGKLITRLASDAPNIKA
H_S09_0015622__-2210 ggtttgacgcacgacaactgaggcaaag
                      accataaccaccgattcgtcgaccatacGTAAGTA  Intron 2
                      ttgtccttctatcgtattgtttagttaa<0-----[2126 : 2061]


CAEEL-PGP-14___   882    VVDARMLQVIYALAAIIANIAIAFIYCWQ
                         VVD R LQVIYA+ A+IA I I FI  WQ
                         VVDGRALQVIYAMTAVIACIIIGFISSWQ
H_S09_0015622__-2063     ggggcgccgatgaaggagtaaagtatatc
                      CAGttaggctattactccttcgtttgttcgga
                      -0>tctttatatccaggtacataatacatcgg


CAEEL-PGP-14___   911                        IGILGTSLILLLAFVMIGLAYKISLM
                                             + ++G  ++++LA  MI LA  I
                                             VTLMGIGMLIVLATSMIWLALTIMNK
H_S09_0015622__-1973                         gacagagatagcgataattgcaaaaa
                      GTAAAGT  Intron 3   AAGtcttgtgtttttcccttgtctcttaa
                      <0-----[1973 : 1918]-0>aaagtaagatcgtatgaggtggcgca


CAEEL-PGP-14___   937 NVEQIQNDDAGRI              -          ---AIEIIEN
                      N+E +++D+AGR+                          AIE IEN
                      NIELVKDDEAGRV              Y          FQIAIETIEN
H_S09_0015622__-1839 aagcgagggggcgt          [tac]       tcagagaaga
                      atattaaaacggt GTTTGTT  Intron 4   GAGatatctactaa
                      ccagcgttataaa <1-----[1799 : 1359]-1>ctgtacagtgt


CAEEL-PGP-14___   957 VKTIQLLTRCELFFDHYQTSSKQQKRSELKK
                      V+TIQLLTR   F+  Y+ +SK  KRSE  K
                      VRTIQLLTRMSTFYGRYKAASKLGKRSESIK
H_S09_0015622__-1326 gcaactcacatattgctaggaacgactgtaa
                      tgctattcgtcctaggaaccgatgagcacta
                      tacaaattagttctaatacctacaaataaca


CAEEL-PGP-14___   988             G          MIEAINYSLTQSFMYFMMCFTYA
                                  G          ++EAIN++++QSF Y+M+C  YA
                                  G          IFEAINFTISQSFTYLMVCVCYA
H_S09_0015622__-1233 g          [gga]       atggaataatcttatcagtgttg
                       GTTCCTT  Intron 5   AAGgttactatctcactcatttgtgac
                       <1-----[1232 : 1181]-1>accacactactacttccgtctttc
```

```
CAEEL-PGP-14___ 1012 VGIRIIYQGDKSSDDTF              K          GIIAMM
                     VGI IIY   K+ D+ F              +           IIAM+
                     VGIHIIYTEQKTPDNVF              R          TIIAML
H_S09_0015622__-1109 ggacaatagcaacgagta          [aga]        aaagac
                     tgtattacaaaccaattgGTAAGAA  Intron 6   CAG cttctt
                     agattctcaggacccac <2-----[1056 :  861]-2>agcatga


CAEEL-PGP-14___ 1036 LGAVAVMNSAQYFPEFVKAKTAAGMLFNIIYRKPRTGDLMEGDR  1079
                     L +VAVMNS+ YFPEFVKA+TAAG+LF++IYRKPRTGD   GD+
                     LASVAVMNSSSYFPEFVKARTAAGLLFSVIYRKPRTGDANVGDK
H_S09_0015622__ -841 cgtgggaatttttcgtgagcagggcctagatcaccagggaggga   710
                     tcctcttacccatcattacgcccgtttgttagacgcgacatgaa
                     cgcgcggtctatccaccaagagaacatcgtctgaagatgttcta


H_S09_0015622__ -709 GTGGTGA     Intron                                    1
                     <0-----[709  :     ]-0>
                            1 n [570  :  570]
//

Gene 1
EXONS 2539 710
  Exon 2539 2462 phase 0
  Exon 2210 2127 phase 0
  Exon 2060 1974 phase 0
  Exon 1917 1800 phase 0
  Exon 1358 1233 phase 1
  Exon 1180 1057 phase 1
  Exon 860 710 phase 2
//

Making a Y in phase 1 intron
Making a G in phase 1 intron
Making a R in phase 2 intron

>H_S09_0015622___HAECO_S09_Supercontig_0015622
_2539-710bp_AA
_CAEEL-PGP-14_828-1079AA

SVFFAIVSENLAMRFRVESFKNLLYQDASYFDNPAHTPGKLITRLASDAPNIKAVVDGRA
LQVIYAMTAVIACIIIGFISSWQVTLMGIGMLIVLATSMIWLALTIMNKNIELVKDDEAG
RVYFQIAIETIENVRTIQLLTRMSTFYGRYKAASKLGKRSESIKGIFEAINFTISQSFTY
LMVCVCYAVGIHIIYTEQKTPDNVFRTIIAMLLASVAVMNSSSYFPEFVKARTAAGLLFS
VIYRKPRTGDANVGDK
//

>H_S09_0015622___HAECO_S09_Supercontig_0015622

agtgtgttcttcgcaatcgtgtcagagaatttggcgatgcgatttcgagtggaatccttc
aaaaatctactctaccaggatgcttcgtatttcgacaatcctgcccatacacctggcaag
cttataactcgtttggctagtgatgcaccgaatattaaagcagttgtcgatggtcgtgca
cttcaagttatctacgcaatgacggctgtaatcgcatgtataataattggattcatatct
agctggcaggtaacactaatgggtataggaatgttaattgtcctggctacatctatgata
tggttggctctgacgatcatgaacaaaaacatcgaactggtcaaggatgatgaagctgga
cgagtatactttcagattgcaatcgaaacgattgagaatgttcgaaccatacaattactt
actcgaatgtctactttctatggacgatataaagccgccagtaaactcggaaaacgatct
gaatcaatcaaaggaatcttcgaagccataaactttacaatctctcaatcctttacttac
ctcatggtttgcgtttgttatgccgtagggatacatattatctataccgaacagaagaca
cccgacaacgtattcagaacgatcatagctatgctactcgcgtccgtggccgtgatgaat
tcctcttcatatttccccgaattcgtcaaagcacggacagcggcaggactcctatttagc
gtgatttaccgtaagccacgaacgggagatgcgaatgttggcgataaa
//
```

A.42 *HAECO*_S09_SUPERCONTIG_0021166 ≈ *CAEEL*-PGP-14

```
genewise
Query protein:       CAEEL-PGP-14___CAEEL-PGP-14 F22E10 3 1-1327AA
Target Sequence      H_S09_0021166___HAECO_S09_Supercontig_0021166

H_S09_0021166__     1          Intron     CAG
                     <1-----[     :  168]-1>


CAEEL-PGP-14___  1030   GIIAMMLGAVAVMNSAQYFPEFVKAKTAAGMLFNIIYRKPRTGDLME
                         IIAM+L +VAVMNS+ YFPEFVKA+TAAG+LF++IYRKPRTGD
                        TIIAMLLASVAVMNSSSYFPEFVKARTAAGLLFSVIYRKPRTGDANV
H_S09_0021166__   169   aaagaccgtgggaatttttcgtgagcagggcctagatcaccagggag
                        cttctttcctcttacccatcattacgcccgtttgttagacgcgacat
                       agcatgatgagcggtctatccaccaagagaatattatctgaaaatgtt


CAEEL-PGP-14___  1077 GDRP                          EIRGNILFENVKFSYPQRPLQP
                      G++                           IRGNILF++VKFSYPQRP QP
                      GEKV                         TIRGNILFDDVKFSYPQRPRQP
H_S09_0021166__   311 ggag                          aacgaactgggatatccccccc
                      gaatGTGAGTT  Intron 1   CAGctggatttaatatgacagcgac
                      tgag<0-----[323  : 1155]-0>cttactgccccgcctgagtagg


CAEEL-PGP-14___  1103 IMKGLQWTALRGQTVALVGPSGSGKSTNIGMLERFYDVTGGAL
                      IM+GLQ++A RGQTVALVGPSGSGKST I MLERFYD TGG +
                      IMRGLQFSAQRGQTVALVGPSGSGKSTIISMLERFYDTTGGYV
H_S09_0021166__  1222 aaagccttgccgcaggcggctgtgataaatacgcttgaaggtg
                      ttggtatccaggactcttgccgcgaccttcttagtaaccggat
                      agagaatatactacaatcaattcagcctaggtatcttttcatt


CAEEL-PGP-14___  1146                          RIDGQDIRKLSLFHLRTQMALVGQEP
                                               R+DG+DI+ LSL HLRTQMALVGQEP
                                               RFDGKDIKTLSLNHLRTQMALVGQEP
H_S09_0021166__  1351                          ctggagaaactcacccacagtggcgc
                      GTGAGTA  Intron 2   CAGgtagaatactctaatgcatcttgaac
                      <0-----[1351 : 1411]-0>actagttgaagcctacgagaataaga


CAEEL-PGP-14___  1172 RLFAGTIRENVCLGLKD                        VPLEKINQA
                      RLF+GTI++N+C GL                          VP+EKI++A
                      RLFSGTIKQNICFGLGV                        VPMEKIDRA
H_S09_0021166__  1490 acttgaaacaattgtgg                         gcagaagcg
                      gttcgctaaatgtgtgtGTAAGGC  Intron 3   AAGtctaatagc
                      ggtgaccagcttccaag<0-----[1541 : 1593]-0>taggaccat


CAEEL-PGP-14___  1198 LELANANRFLANLPAGIDTD                        1217
                      LELANA  FLANLPAGIDT+
                      LELANAKNFLANLPAGIDTE
H_S09_0021166__  1621 cgtgagaatcgatcggagag                        1680
                      tatcacaattcatccgtaca
                      cgagtcatttttaactccag


H_S09_0021166__  1681 GTAAGGA  Intron                             1759
                      <0----[1681 :     ]-0>
//

Gene 1
EXONS 169 1680
  Exon 169 322 phase 0
  Exon 1156 1350 phase 0
  Exon 1412 1540 phase 0
  Exon 1594 1680 phase 0
//

>H_S09_0021166___HAECO_S09_Supercontig_0021166
_169-1680bp_AA
_CAEEL-PGP-14_1030-1217AA

TIIAMLLASVAVMNSSSYFPEFVKARTAAGLLFSVIYRKPRTGDANVGEKVTIRGNIL
FDDVKFSYPQRPRQPIMRGLQFSAQRGQTVALVGPSGSGKSTIISMLERFYDTTGGYVRF
DGKDIKTLSLNHLRTQMALVGQEPRLFSGTIKQNICFGLGVVPMEKIDRALELANAKNFL
ANLPAGIDTE
//

>H_S09_0021166___HAECO_S09_Supercontig_0021166

AACGATCATAGCTATGCTACTTGCGTCAGTGGCCGTGATGAATTCCTCTTCATAT
TTCCCCGAATTCGTCAAAGCACGGACAGCGGCAGGACTTCTATTTAGTGTAATTTACCGT
AAGCCACGAACAGGAGATGCGAATGTTGGTGAGAAAGTGACCATTCGTGGAAACATTCTG
TTCGACGACGTCAAGTTCAGCTATCCGCAACGGCCTCGACAGCCGATAATGAGAGGGCTA
CAATTTTCAGCTCAACGCGGTCAAACCGTAGCACTTGTCGGACCATCTGGTTCCGGAAAG
TCCACCATTATATCGATGCTTGAACGTTTCTATGATACTACTGGCGGATATGTTCGATTC
GATGGAAAGGATATTAAGACACTATCGCTCAACCATCTACGCACGCAAATGGCATTAGTT
GGACAAGAGCCAAGGCTGTTTTCGGGAACCATCAAACAGAACATTTGTTTCGGCTTAGGA
GTGGTTCCAATGGAGAAAATCGACCGAGCTCTCGAGTTAGCGAATGCCAAAAATTTTCTT
GCTAATTTACCAGCCGGTATCGACACAGAG
//
```

A.43 *HAECO*_S09_SUPERCONTIG_0010288 ≈ *CAEEL*-PGP-13

```
genewise
Query protein:      CAEEL-PGP-13___CAEEL-PGP-13 F22E10 2 1-1324AA
Target Sequence     H_S09_0010288___HAECO_S09_Supercontig_0010288

H_S09_0010288__-1199                    ACGA                    1196
                     <?-----[     : 1196]-?>


CAEEL-PGP-13___ 1224 GGQKQRIAIARALVRDPKILLLDEATSALDSESERA
                     GGQKQRIAIARALVRDPKILLLDEATSALDSESER
                     GGQKQRIAIARALVRDPKILLLDEATSALDSESERV
H_S09_0010288__-1195 ggcaccagagcgcgcgcaatccgggaagtgtgagag
                     ggaaagtctcgcttgacattttaaccgctacagagt
                     tagggtcctaggatatcgagggcacctcgtaataaa


CAEEL-PGP-13___ 1260                  V          QEALDRAREGRTCITIAHRLSSI
                                      +          Q+ALD AREGRTCITIAHRLSSI
                                      I          QKALDLAREGRTCITIAHRLSSI
H_S09_0010288__-1087 a            [atc]          cagcgtgcggcataaagccctta
                      GTTGAAT  Intron 1   CAGtaactatcgaggcgtctcagtcct
                      <1-----[1086 :  268]-1>caacgtgtaattactacctaagat


CAEEL-PGP-13___ 1284 QNSDLIVYIDDGRVQESGTHKELMQLKGKYFELIKKQDLAI     1324
                     QN+DLIVY+++G+V+ESGTH +LMQ +G Y++LIKKQDL
                     QNADLIVYVENGKVRESGTHSQLMQRRGCYYQLIKKQDLTT
H_S09_0010288__ -196 caggcagtggagagcgtgactctacccgtttccaaacgcaa       74
                     aacatttataagatgacgcacattaggggaaattaaaatcc
                     attctcctgacggtgatcttcggggattcctagcggattga


CAEEL-PGP-13___       *
                     *
                     *
```

```
H_S09_0010288__  -73 t                                                71
                     g
                     a


H_S09_0010288__  -70 [70   :    1]                                     1
//

Gene 1
EXONS 1195 74
  Exon 1195 1087 phase 0
  Exon 267 74 phase 1
//

Making a I in phase 1 intron


>H_S09_0010288___HAECO_S09_Supercontig_0010288
_1195-74bp_AA
_CAEEL-PGP-13_1224-1324AA_End

GGQKQRIAIARALVRDPKILLLDEATSALDSESERVIQKALDLAREGRTCITIAHRLSSI
QNADLIVYVENGKVRESGTHSQLMQRRGCYYQLIKKQDL
TT
//

>H_S09_0010288___HAECO_S09_Supercontig_0010288
ggtggacagaagcagcgtatcgccattgcacgggcgctagttcgagatcccaagatattg
ctgctggacgaagccaccagtgccttggattcagaaagtgaaagagtaatccaaaaagcc
ctggatttggctcgagaaggtcgtacatgcattacaatcgcccatcgactatcgtcaatt
caaaatgctgaccttatcgtctatgtggaaaacgggaaggttcgggaatctggcactcat
tcccagttgatgcagcgacgtggttgctactatcaactgatcaagaagcaagatctt
acgaca
//
```

A.44 *HAECO*_S09_SUPERCONTIG_0024244 ≈ *CAEEL*-PGP-14/15

```
genewise
Query protein:       CAEEL-PGP-14___CAEEL-PGP-14_F22E10_3_1-1327AA
Comp Matrix:         blosum62.bla
Gap open:            12
Gap extension:       2
Start/End            local
Target Sequence      H_S09_0024244___HAECO_S09_Supercontig_0024244
Strand:              both
Start/End (protein)  local
Gene Paras:          worm.gf
Codon Table:         codon.table
Subs error:          1e-05
Indel error:         1e-05
Model splice?        model
Model codon bias?    flat
Model intron bias?   tied
Null model           syn
Algorithm            623

H_S09_0024244__     1          Intron     TAG
                     <2-----[     : 1026]-2>


CAEEL-PGP-14___   702  SGSEAFRRGNSLNDSFSGSKRSAQADAENSAFAANEAAIMAE
                       SGSEAF RG +LNDSF     +A+ADAEN A+A     IM E
                       SGSEAFVRGQALNDSFGRQSYNAEADAENEALALEVKKIMEE
H_S09_0024244__  1027  tgtggtgcgcgcagttgccttaggggggaggcgcggaaaagg
                       cgcacttggactaactggacaacacacaaactctataattaa
                     atctagtctggtttcgtgaaatttaacaatacttggggacgag


CAEEL-PGP-14___   744                        DGQITAGYLDIFKNAKGNYLYMFLGT
                                            DG I+AGY+DI+KNA GNY ++FLG
                                            DGVISAGYIDIYKNATGNYHWIFLGF
H_S09_0024244__  1154                       gggaaggtagataagagatctatcgt
                     GTAAGGA  Intron 1   TAGagttgcgatataaaccgaaagtttgt
                     <0-----[1154 : 1692]-0>ttctttaccccccaccaatccgactct


CAEEL-PGP-14___   770 VFALIR                                          775
                       V A++R
                       VTAVFR
H_S09_0024244__  1771 gaggtc                                        1788
                       tccttg
                       cagtct
                              1683 n [1789 : 3471]
                                 1 n [3527 : 3527]


CAEEL-PGP-14___   776 GLELPALALIFGWVFEGFTFVPYGGRMMHRMAMAVIAFASVGVGVWFSQL
H_S09_0024244__       ------------------------------------------------


CAEEL-PGP-14___   826 ASSVLFAVVSENLSMRFRVQSFRNLLYQ                    853
H_S09_0024244__       ----------------------------


H_S09_0024244__          Intron     TAG                            3895
                     <0-----[     : 3895]-0>


CAEEL-PGP-14___   854 DASYFDNPAHAPGKLITRLASDAPNIKA
                       DASYFDNPAH PGKLITRLASDAPNIKA
                       DASYFDNPAHTPGKLITRLASDAPNIKA
H_S09_0024244__  3896 ggtttgacgcacgacaactgaggcaaag
                     accataaccaccgattcgtcgaccatacGTAAGTT  Intron 1
                     ttgtccttctatcgcattgttcagttaa<0-----[3980 : 4036]


CAEEL-PGP-14___   882    VVDARMLQVIYALAAIIANIAIAFIYCWQ
                         VVD R LQVIYA+ A+IA I I FI  WQ
                         VVDGRALQVIYAMTAVIACIIIGFISSWQ
H_S09_0024244__  4034    ggggcgccgatgaaggagtaaagtatatc
                     CAGttaggctattactccttcgtttgttcgga
                     -0>tttttatatccaggtacataatacaccgg


CAEEL-PGP-14___   911                        IGILGTSLILLLAFVMIGLAYKISLM
                                            + ++G  ++++LA  MI LA  I
                                            VTLMGIGMLIILATSMIWLALTIMNK
H_S09_0024244__  4124                       gacagagataacgataattgtaaaaa
                     GTAAAAT  Intron 2   TAGtcttgtgtttttcccttgtctcttaa
                     <0-----[4124 : 4178]-0>aaagtaggatcgtgtgagatagcgta


CAEEL-PGP-14___   937 NVEQIQNDDAGRI                                  949
                       N+E +++D+AGR+
                       NIELVKDDEAGRV
H_S09_0024244__  4257 aagcgaggggggcgt                               4296
                       atattaaaacggt
                       ccagcgttataaa


H_S09_0024244__  4297  GTTTATT  Intron 3
                       <1-----[4297 :     ]-1>
                       1337 n [4735 : 6071]


CAEEL-PGP-14___   950 AIEIIENVKTIQLLTRCELFFDHYQTSSKQQKRSELKKGMIEAINYSLTQ
H_S09_0024244__       ------------------------------------------------


CAEEL-PGP-14___  1000 SFMYFMMCFTYAVGIRIIYQGDKSSDDTFKGIIAMMLGAVAVMNSAQYFP
H_S09_0024244__       ------------------------------------------------


CAEEL-PGP-14___  1050 EFVKAKTAAGMLFNIIYRKPRTGDLMEGDRP                 1080
H_S09_0024244__       --------------------------------


H_S09_0024244__          Intron     CAG                            6380
                     <0-----[     : 6380]-0>


CAEEL-PGP-14___  1081 EI
                       I
                       TI
H_S09_0024244__  6381 aa
                       ct
                       ct


CAEEL-PGP-14___  1083 RGNILFENVKFSYPQRPLQPIMKGLQWTALRGQTVALVGPSGSGKSTNI
                       RGNILF++VKFSYPQRP QPIM+GLQ++A RGQTVALVGPSGSGKST I
                       RGNILFDDVKFSYPQRPRQPIMRGLQFSAQRGQTVALVGPSGSGKSTII
H_S09_0024244__  6387 cgaactgggatatccccccccaacgccttgccgcaggcggctgtgataaa
                                                    ggatttaatatgacagcgacttggtatccaggactcttgccgcgacctt
                                                    tactgccccgcctgagtgatagatgacatactattatcaattcagtctc


CAEEL-PGP-14___  1132 GMLERFYDVTGGAL                                RIDGQDIRKLSL
                       MLERFYD TGG +                                R+DG+DI+ LSL
                       SMLERFYDTTGGYV                               RFDGKDIKTLSL
H_S09_0024244__  6534 tacgcttgaaggtg                                ctggagaaactc
                     cttagtaaccggatGTGAGTA  Intron 5   CAGgtagaatactct
                     agtgttttctcatt<0-----[6576 : 6636]-0>actagttgaagc


CAEEL-PGP-14___  1158 FHLRTQMALVGQEPRLFAGTIRENVCLGLKD
                       HLRTQMALVGQEPRLF+GTI++N+C GL
                       NHLRTQMALVGQEPRLFSGTIKQNICFGLGV
H_S09_0024244__  6673 acccacagtggcgcacttgaaacaattgtgg
                     aatgcatcttgaacgttcgctaaatgtgtgt
                     ctacgagaataagagatgagtagcttccaag


CAEEL-PGP-14___  1189                        VPLEKINQALELANANRFLANLPAGI
                                             VP+EKI++ALELANA  FLANLPAGI
                                             VPMEKIDRALELANAKGFLANLPAGI
H_S09_0024244__  6766                        gcagaagcgcgtgagagtcgatcgga
                     GTGAGGC  Intron 6   AAGtctaatagctatcacagttcatccgt
                     <0-----[6766 : 6818]-0>aaggaccggtgacccatctttaactc


CAEEL-PGP-14___  1215 DTD                           VGEKGGQLSGGQKQRIAIARALV
                       DT+                           VGEKG QLSGGQKQRIAIARALV
                       DTE                           VGEKGTQLSGGQKQRIAIARALV
H_S09_0024244__  6897 gag                           gggagacctggcaccagagcgcg
                       acaGTAATGA  Intron 7   CAGtgaagcatcggaaagtctcgctt
                       cgg<0-----[6906 : 8184]-0>ctaacaacgtagggtcctagaat


CAEEL-PGP-14___  1241 RDPKILLLDEATSALDSESER                         1261
                       RDPKILLLDEATSALDSESER
                       RDPKILLLDEATSALDSESER
H_S09_0024244__  8254 cgcaatccgggaagtgtgaga                         8316
                       gacatttaaccgctacagag
                       atcgagggcacctcgtaataa


H_S09_0024244__  8317 GTAAGTT  Intron 8                             8633
                       <0-----[8317 :     ]-0>
//

Gene 1
EXONS 1027 8316
  Exon 1027 1153 phase 0
  Exon 1693 1788 phase 0

  ---

  Exon 3896 3979 phase 0
  Exon 4037 4123 phase 0
  Exon 4179 4296 phase 0

  ---

  Exon 6381 6575 phase 0
  Exon 6637 6765 phase 0
  Exon 6819 6905 phase 0
  Exon 8185 8316 phase 0
//

>H_S09_0024244___HAECO_S09_Supercontig_0024244
_1027-1788bp_AA
---
_3896-4296bp_AA
---
_6381-8316bp_AA
_CAEEL-PGP-14_OR_15_CAEEL-PGP-14_702-775AA---854-949AA---1081-1261AA

SGSEAFVRGQALNDSFGRQSYNAEADAENEALALEVKKIMEEDGVISAGYIDIYKNATG
NYHWIFLGFVTAVFR

DASYFDNPAHTPGKLITRLASDAPNIKAVVDGRALQVIYAMTAVIACIIIGFISSWQVTL
MGIGMLIILATSMIWLALTIMNKNIELVKDDEAGRV

TIRGNILFDDVKFSYPQ
RPRQPIMRGLQFSAQRGQTVALVGPSGSGKSTIISMLERFYDTTGGYVRFDGKDIKTLSL
NHLRTQMALVGQEPRLFSGTIKQNICFGLGVVPMEKIDRALELANAKGFLANLPAGIDTE
VGEKGTQLSGGQKQRIAIARALVRDPKILLLDEATSALDSESER
//

>H_S09_0024244___HAECO_S09_Supercontig_0024244
 1027-1788bp
---
_3896-4296bp
---
_6381-8316bp

atctggctctgaagcgtttgtccgtgggcaggctcttaatgactcgtttgggcgacaa
tcatataatgctgaagcagacgcagaaaatgaagccttgctctggaggtgaagaaaatc
atggaagaggatggtgtcattagtgctggatacatcgacatctacaaaaacgccacagga
aattaccactggatattccttggctttgtcacagcggtttccgt


gatgcttcgtatttcgacaatcctgcccatacacctggcaagctcataactcgtttggct
agtgacgcaccgaatattaaagcagttgttgatggtcgtcgacttcaagttatctacgca
atgacggctgtaatcgcatgtatataattggattcatatccagctggcaggtaacacta
atgggtatagggatgttaattatcctggctacgtctatgatatggttagctttaacgatc
atgaataaaaacatcgaactggtcaaggatgatgaagctggacgagtat
accattcgtggaaacattctgttcgacgacgtcaagttcagctat
ccgcaacggcctcggcaacctataatgcgaggtctgcaattctcagctcaacgcggtcaa
actgttgcacttgtcggaccatctggttccggaaagtctaccattatctcaatgcttgag
cgtttttatgataccactggcggatatgttcgattcgatggaaaggatattaagacacta
tcgctcaaccatctacgcacgcaaatggcattagttggacaagagccaaggctattttcg
ggaacgattaaacagaacatttgtttcggcttaggagtggtaccaatggagaaaatcgac
cgggcgcttgagttagccaacgccaaaggtttccttgctaatttaccagccggtatcgac
acggaggtcggtgaaaaggcacacaactctcgggtggacagaagcagcgtatcgccatt
gcacgggcactagttcgagatcccaagatattgctgctggacgaagccaccagtgccttg
gattcagaaagtgaaaga
```

A.45 *HAECO*_S09_SUPERCONTIG_0005706 ≈ *CAEEL*-PGP-15PS

```
genewise
Query protein:        CAEEL-pgp-15ps_F22E10_4_not_counting_X_1-1320AA
Comp Matrix:          blosum62.bla
Gap open:             12
Gap extension:        2
Start/End             local
Target Sequence       H_S09_0005706___HAECO_S09_Supercontig_0005706
Strand:               both
Start/End (protein)   local
Gene Paras:           worm.gf
Codon Table:          codon.table
Subs error:           1e-05
Indel error:          1e-05
Model splice?         model
Model codon bias?     flat
Model intron bias?    tied
Null model            syn
Algorithm             623

H_S09_0005706__     1          Intron      TAAG
                    <?-----[1    :   56]--2>


CAEEL-pgp-15ps_   144  NVYIFLGIGIFVTITNYIQYMCFQHCCSRIMAQMRHRYVYSVLRQNAG
                         +++F G  +  ++ NY+   +    +     I+ ++R  +V +VL  QNA
                        FLFLFYGRVVSKSLQNYFLSLASHN----IVGRIRKEFVKAVLAQNAA
H_S09_0005706__     57  ttttttgcggtaaccattcttgaca    aggcacagtgaggcgcagg
                        attttaggttcagtaaattctcgaa    ttggtgaattacttcaacc
                        gtgtactatctcattgtttcggtttt    ttacacgacgaaagagccg


CAEEL-pgp-15ps_   192 WFDKHHSGTIATKLN              D          SMERIREG
                        WFD++++GTI TKLN              +          ++ +I +G
                        WFDENNAGTITTKLN              E          NVAQIEDG
H_S09_0005706__   191 ttggaaggaaaaacag          [gag]        aggcaggg
                        gtaaaacgctccataaGTAGGCA  Intron 1   CAG atcataag
                        gtcaccaagccgggc <2-----[238  :  681]-2>gcatatgta


CAEEL-pgp-15ps_   216 IGDKLGVLLRGVAMLVASVVVAYIYEWRLACMLLGVAPTCIGC
                        IGDK+G+L RGV + +AS   A+ Y+WR+  + +   P
                        IGDKIGMLARGVTVFIASAAFAFYYDWRITLVCIWDGPVSAIT
H_S09_0005706__   707 aggaagacgaggagtagaggtgtttgtcaacgtatggcgagaa
                        tgaatgttcggtctttcgcctctaaaggtcttgtgagctgctc
                        ttcgcaggaaagagctccctttttccgtctggctgctttccat


CAEEL-pgp-15ps_   259 MSLMAR                                          264
                        M++M+R
                        MAIMSR
H_S09_0005706__   836 agaata                                         853
                        tcttcg
                        gccgag


CAEEL-pgp-15ps_   265 QMTATTVKEL              G          GVEK        279
                                +                        E
                        VSENLENSQF             F         SGE-
H_S09_0005706__   854 gagacgatctt          [ttt]       tgg        1170
                        tgaataacat GTAGACA  Intron 2   CAGtcga
                        acgtgactac <1-----[885  : 1159]-1>ttta


CAEEL-pgp-15ps_   280 AGSIAEESLMGVRTVQAFNGQEEMVE
                        AG+IAEE++M V+TV A NGQ+ MV+
                        AGAIAEEAIMNVKTVAACNGQKHMVK
H_S09_0005706__  1171 gggagggggaaagaagggtagcacaga
                        cgctcaacttatactccgagaaattaGTGGGTT  Intron 3   CAG
                        ttgataactgccaacatccaagtgta<0-----[1249 : 3651]-0>
                                                    1 n [2720 : 2720]
                                                    1 n [2726 : 2726]
                                                   10 n [2843 : 2852]


CAEEL-pgp-15ps_   306 RYRVELNKGRKFAIWKGFWSGLFGGMFFFWLFAFQGCGF
                        +Y  +L +G  FAI   F +G   G++FF L+ F    F
                        KYDEQLRRGVSFAIRYSFINGFCEGFMFFQLYIFYAAAF
H_S09_0005706__  3652 atggccaaggttgactataagttggtattcctattgggt
                        aaaaatgggtctctgagttagtgagttttatattaccct
                        attgagaaaagtttatccctatcatcgctacttccgttc


CAEEL-pgp-15ps_   345                L          YGAYLLSVGIIKSPGDVFIIVMA
                                    L          YG      GI    PG +FI+
                                    L          YGIPSYYNGITAEPGTIFIVAST
H_S09_0005706__  3769 c          [cta]        tgacattagaaggcgaataggta
                        tGTGAGTC  Intron 4   TAG agtcgaaagtccacgcttttccc
                         <2-----[3771 : 3850]-2>atatacttttcttatagtttgcta


CAEEL-pgp-15ps_   369 MLL              G          SYFTGLISPHLMVLLNARVA
                        +LL              G          SYF GL+ PH+M ++ AR+A
                        VLL              G          SYFFGLLGPHMMAIMKARIA
H_S09_0005706__  3921 gctg           [ggc]        ttttgccgccaagaaagaag
                        ttt GTGAGGT  Intron 5   TAGgcattgttgcattcttacgtc
                        tga <1-----[3931 : 4006]-1>cccctagccttggctggaaat


CAEEL-pgp-15ps_   393 AGSIYETIDRVPKID                          -PYSKKGKFLD
                        A  IYETID+V K+                             SK+GK L
                        AAIIYETIDQVNKFP                          DIVSKEGKELR
H_S09_0005706__  4069 ggaatgaagcgaatc                           gagtaggagta
                        ccttaactaataatcGTCCTGT  Intron 6   CAGattcaagaatg
                        ggcctagttaatata<0-----[4114 : 4341]-0>tctcaaaaagg


CAEEL-pgp-15ps_   418 KVIGRVKFENVHFRYPT              R          KDAKIL
                          GR++F +VHF+YPT              R            IL
                        ACKGRLEFRDVHFKYPT              R          -ETPIL
H_S09_0005706__  4375 gtagctgtaggctatcaa          [agg]        gacac
                        cgaggtatgataaccgGTGACAT  Intron 7   TAG  acctt
                        gcagagacgcttcatac <2-----[4428 : 4482]-2>g ggtag


CAEEL-pgp-15ps_   442 NGLNLTIEPGTSVALVGHSGCGKSTS
                         GL+   EPG ++A VG SGCGKSTS
                        QGLSWVAEPGETIAFVGKSGCGKSTS
H_S09_0005706__  4499 cgcatgggcggaagtggaagtgaaaa
                        agtggtcacgactcttgaggggagcgGTTCGTT  Intron 8   CAG
                        atctgagataggtgcatacctcacct<0-----[4577 : 4665]-0>


CAEEL-pgp-15ps_   468 VGLLTRLYEPEAGNVTIDGTDVRELNIDYLRN
                        +GLLTRLY+ + G+  +DG ++R +    LR
                        IGLLTRLYDCDKGSALLDGQEIRSIKTSDLRK
H_S09_0005706__  4666 agccacctgtgagtgccggcgactaaaagtaa
                        tgttcgtaagaagccttagaatgctacgatga
                        ccagtgtcctcaaatctttggtagcagttgag


CAEEL-pgp-15ps_   500                        VIGIVQQEPILFNDTIHNNLLLGNPN
                                             +IGIVQQEP LFN TI  N++LG
                                            MIGIVQQEPCLFNGTIRENIVLGR-S
H_S09_0005706__  4762                       aagagccgctctagaacgaagcgc t
                        GTTCTAT  Intron 9   CAGttgttaaacgttagctgaatttgg c
                        <0-----[4762 : 4817]-0>gtctcgagtttttctttgcagtca a


CAEEL-pgp-15ps_   526 ATREKMIEVCKMANAHDFIEKMPK                        GY
                         T E+  +  ++ANAHDFI K+ K                      GY
                        ITDEQAEDAARIANAHDFIMKLDK                       GY
H_S09_0005706__  4893 aaggcgggggcagagcgtaaacga                         gt
                        tcaaacaaccgtcacaatttataaGTGAGCT  Intron 10  AAGga
                        ctcaacatttaagttctcagagca<0-----[4965 : 5042]-0>cc


CAEEL-pgp-15ps_   552 DTMIGDGGVQLSGGQKQRVAIARTLVREPKVLLLDEATSALDAQSESIV
                        DT+IG GGV LSGGQKQR+AIAR +  +PK+LLLDEATSALD++SE++V
                        DTIIGSGGVSLSGGQKQRIAIARAVATQPKILLLDEATSALDSESENVV
H_S09_0005706__  5049 gaaagtgggtctggcacaagagcgggaccaatctgggaagcgtgagagg
                        acttgcggtctcggaaagtctcgctccacattttaaccgctacagaatt
                        ccttatcgcatgaagaggttttaggatggatgtgcgtctagttacatgg


CAEEL-pgp-15ps_   601 Q                        SALNNASKGRTTIMIAHRLSTIREA
                      Q                         ALN AS+GRTTI+IAHRLST+++
                      Q                        LALNRASRGRTTIVIAHRLSTLKDV
H_S09_0005706__  5196 c                        cgcaagtcgcaaagagcctaatagg
                      aGTATGAA  Intron 11  TAGtctagccgggcctttcagtgctaat
                      g<0-----[5199 : 5266]-0>ttgcgatttcagacttccgcgggtc


CAEEL-pgp-15ps_   627 DKIVFFEKGVIVEA                                  640
                       +I  ++ G +VEA
                      QRIYAIQDGKVVEA
H_S09_0005706__  5342 ccatgacggaggggg                                5384
                      agtactaagattac
                      ggttccacgaagaa


H_S09_0005706__  5385 GTATTAC  Intron
                      <1-----[5385 :     ]-1>
                        10 n [6078 : 6087]
                         1 n [6165 : 6165]


CAEEL-pgp-15ps_   641 GNHEELVRLGGRYYDLVKAQAFKPDNDGTVDEIEDEIDLGGLHSRQSSFT
H_S09_0005706__       --------------------------------------------------


CAEEL-pgp-15ps_   691 SSIRSKMSGAEAFRRGTLGADSFAGGKSTARADAENAAFAEEVAKVMEQD
H_S09_0005706__       --------------------------------------------------


CAEEL-pgp-15ps_   741 GQISAGYMDIFKNAKGNYTVMFLGFVTGLIRGLELTAFALLLGWVFEGFQ
H_S09_0005706__       --------------------------------------------------


CAEEL-pgp-15ps_   791 YLNIDNGKMMHRMAMAVIAYGCSGFGCFISQFLSSIFFAIVSENLALRFR
H_S09_0005706__       --------------------------------------------------


CAEEL-pgp-15ps_   841 VMSFRNLLYQDASFFDNPAHAPGKLITRLATDAPNCKTVVDSRMLQVLYA
H_S09_0005706__       --------------------------------------------------


CAEEL-pgp-15ps_   891 LSAIIANIVISFIYCXILGTALIILLAFTMCGLAYKISLLNMEQIQNDEA
H_S09_0005706__       --------------------------------------------------


CAEEL-pgp-15ps_   941 GR                                              942
H_S09_0005706__       --


H_S09_0005706__            Intron     CAG
                      <0-----[      : 8757]-0>


CAEEL-pgp-15ps_   943  IAIEIIENVKTIQLLTRCEHFFDHYEKCSKSQKRSELK
                       ++IE++E+ KTIQLL   ++F+  YE      K+ E
                       LSIEVFEHAKTIQLLAVQDYFLQKYESYEAVVKKQEKW
H_S09_0005706__  8758  ctaggtgcgaaacctggcgtttcatgatggggaacgat
                       tctattaacactattctaaattaaaagaacttaaaaag
                       cctagcgccatcaggctgtctgaatatcgctcgaagag


CAEEL-pgp-15ps_   981                          KGLIEAINYTITQSFMYYMMCFCFAL
                                                + ++I + +TQS++Y+     + +
                                               TTIYQSIQFGLTQSYIYFSDLVTYGI
H_S09_0005706__  8872                          aaatctactgcacttatttgcgatga
                      GTATGTT  Intron      CAGcctaactatgtcacatatcattcagt
                      <0-----[8872 : 9029]-0>agatggcgtgccgtccttctgaccat


CAEEL-pgp-15ps_  1007 GIRLIYHGNKSPQD                                 1020
                      G  +IY G     +D
                      GASMIYFGRVDSKD
H_S09_0005706__  9108 ggaaattgcggtag                                 9149
                      gcgttatggtacaa
                      atcgtctatttggc


H_S09_0005706__  9150  ACAGTCG  Intron                               9807
                       <?-----[     : 9807]-?>

//

Gene 1
EXONS 57 9149

  Exon 57 237 phase 0
  Exon 682 853
       854 884 phase 2
  Exon 1160 1170
       1171 1248 phase 1
  Exon 3652 3770 phase 0
  Exon 3851 3930 phase 2
  Exon 4007 4113 phase 1
  Exon 4342 4427 phase 0
  Exon 4483 4576 phase 2
  Exon 4666 4761 phase 0
  Exon 4818 4964 phase 0
  Exon 5043 5198 phase 0
  Exon 5267 5384 phase 0

  ---

  Exon 8758 8871 phase 0
  Exon 9030 9149 phase 0
//

Making a E in phase 2 intron

Making a F in phase 1 intron

Making a L in phase 2 intron
Making a G in phase 1 intron
Making a R in phase 2 intron


>H_S09_0005706___HAECO_S09_Supercontig_0005706
_57-853bp_AA
_854-1170bp_AA
_1171-5384bp_AA
---
_8758-9149bp_AA
_CAEEL-pgp-15ps_144-264AA_265-279AA_280-640AA---943-1020AA

FLFLFYGRVVSKSLQNYFLSLASHNIVGRIRKEFVKAVLAQNAAWFDENNAGTITTKLN
ENVAQIEDGIGDKIGMLARGVTVFIASAAFAFYYDWRITLVCIWDGPVSAITMAIMSR

VSENLENSQFFSGE

AGAIAEEAIMNVKTVAACNGQKHMVKKYDEQLRRGVSFAIRYSFINGF
CEGFMFFQLYIFYAAAFLYGIPSYYNGITAEPGTIFIVASTVLLGSYFFGLLGPHMMAIM
KARIAAAIIYETIDQVNKFPDIVSKEGKELRACKGRLEFRDVHFKYPTRETPILQGLSWV
AEPGETIAFVGKSGCGKSTSIGLLTRLYDCDKGSALLDGQEIRSIKTSDLRKMIGIVQQE
PCLFNGTIRENIVLGRSITDEQAEDAARIANAHDFIMKLDKGYDTIIGSGGVSLSGGQKQ
RIAIARAVATQPKILLLDEATSALDSESENVVQLALNRASRGRTTIVIAHRLSTLKDVQR
```

IYAIQDGKVVEA

LSIEVFEHAKTIQLLAVQDYFLQKYESYEAVVKKQEKWTTIYQSIQFGLTQSYIYFSDL
VTYGIGASMIYFGRVDSKD
//

>H_S09_0005706___HAECO_S09_Supercontig_0005706
_57-853bp
_854-1170bp
_1171-5384bp

_8758-9149bp

agtttttgtttttattctatggacgtgtcgtttccaaaagtcttcagaattattttctc
tcgttggctagtcataatattgttggacgcatacgcaaggaattcgtgaaagcagtactg
gcacagaacgccgcgtggtttgacgaaaacaacgcaggaacgatcaccacgaagctgaac
gagaacgtagctcaaattgaggatggaattggtgacaagatcggaatgctggcaagagga
gtgacagtgttcattgccagcgccgcttttgctttttactacgactggcgtatcactctg
gtgtgcatttgggacggtcctgttagcgccataactatggccatcatgtcaagg

gtaagc
gagaatctggaaaactctcaattcttttctggtgaa

gctggtgcgatagctgaagaagcc
attatgaacgtcaaaacagtcgcagcttgcaacggacaaaagcatatggttaaaaaatat
gatgagcaactgagaagaggagtatcgtttgctattcgatatagcttcatcaatggattt
tgcgaaggtttcatgttctttcaactctatattttctacgcggctgctttcctatatgga
attccaagctattataatggtatcactgctgaacctggaacgatttttattgtggcctct
acagttctgttaggctcctacttctttggactgctcggccctcatatgatggccattatg
aaggcaagaatagctgcggcgatcatctatgaaacgattgatcaagtaaataaatttcca
gatatcgtttccaaagaaggaaagaattgagggcgtgcaaagggcgattggaattcagg
gacgttcatttcaaatatccaaccagggagacgcctatactgcaaggtctcagttgggta
gcggaacctggagagacgattgcgttcgtaggtaaaagcggctgtggcaaaagcaccagt
atcggcctactgactcggctttacgactgtgacaaaggatcagctctccttgatggtcag
gagattcgatcgatcaaaacgagtgatttgagaaagatgattggcattgtccagcaagag
ccttgtctttttaatggcactattcgtgagaacatagtgcttggccgatcaatcactgac
gaacaagccgaagatgctgctcgaatagcgaatgctcacgatttcataatgaaactggac
aaaggctacgacaccattattggatctggcggggtctcactttcgggaggacagaaacag

aggattgctattgctcgagcggtggcaactcagccgaaaattttgcttttggacgaggct
accagtgcactggattctgaaagcgaaaatgtggtgcagcttgctctgaacagggcatct
cgtggtcgcacaacgatagtcattgctcaccgcttgagcacgttgaaggatgtccagcgg
atttatgccatccaagacgggaaagtagtggaagcag

ctctccattgaagtgttcgagcacgccaaaactatccaactgttggccgttcaggat
tacttttgcaaaaatatgaaagttacgaggccgttgtcaagaaacaagagaaatggaca
acgatatatcagtcgatccagtttgggctcacccagtcttacatctatttttccgatctg
gtaacctacggaattggagctagcatgatttactttggacgtgttgattcgaaggac

A.46 *HAECO*_S09_SUPERCONTIG_0013014 ≈ *CAEEL*-PGP-1/5/6

```
genewise
Query protein:      CAEEL-PGP-5_1a CAEEL-PGP-5 C05A9 1a 1-1252AA
Target Sequence     H_S09_0013014___HAECO_S09_Supercontig_0013014

H_S09_0013014_-35867                ACTATGT                              5437
                    <?-----[     : 5437]-?>
                      10 n [6322 : 6313]
                      10 n [8150 : 8141]
                     602 n [12184:11583]
                      10 n [15789:15780]
                      55 n [18214:18160]
                      76 n [20237:20162]
                      10 n [21487:21478]
                      10 n [26637:26628]
                      10 n [31508:31499]


CAEEL-PGP-5_1a_ 1003 GEIELKNVSFEYAQRSDKMILDGVSLKLPAGRTLALVGPSGS
                     GE+  +NVSF Y  R   ++ + ++L +PAG+ +AL GPSG
                     GEVRFENVSFAYPTRPHHQVFECLNLTIPAGQVVALCGPSGE
H_S09_0013014__-5438 gggctgagatgtcaaccccgtgttacaacggcgggctgcagg
                     gatgtaatgtcaccgcaaattagtatctccgattctggcgga
                     aattcatactatgtgaccgccacgctaaactactaccaacga


CAEEL-PGP-5_1a_ 1045                 G          KSTIISLLERFYHAVDGEVKIDE
                                    G          KSTI SLLERFY  + G V +D
                                    G          KSTITSLLERFYTPLSGRVLLDN
H_S09_0013014__-5312 g          [ggt]        aaaaatccgcttacctgcgctga
                      GTACGCT  Intron 1   TAGgagctccttagtacctcggtttaa
                      <1-----[5311 : 5232]-1>tatgatagtagctgactaagtgct


CAEEL-PGP-5_1a_ 1069 ENV                                              1071
                     +++
                     QDL
H_S09_0013014__-5160 cgc                                             5152
                     aat
                     atg


CAEEL-PGP-5_1a_ 1072 V                                               1072

                     !!
H_S09_0013014__-5151 ca                                              5150


CAEEL-PGP-5_1a_ 1073 DVNLHHLRES                          -VSLVSQEPVLF
                      +NL  LR                             + L+SQEPVLF
                     TLNLEWLRGQ                         VIGLISQEPVLF
H_S09_0013014__-5149 acatgttcgc                          gagtatcgcgct
                     ctatagtggaGTATGTT  Intron 2   GAGttgttcaacttt
                     ggtggggagg<0-----[5119 : 4973]-0>ctagtgaaattc


CAEEL-PGP-5_1a_ 1094 NCSIKENFLFGISHNASQLE                          IDQALK
                       SI+EN+ +G   +A+  E                          + QA +
                     ATSIEENIRYG-KPSATDEE                        VMQAAR
H_S09_0013014__-4936 gaaaggaactg actgaggg                        gacgga
                     ccgtaaatgag accccaaaGTGGTTA  Intron 3   CAGttaccg
                     ttttggtttct agatatgg<0-----[4879 : 3842]-0>ggattg
                      10 n [4385 : 4376]


CAEEL-PGP-5_1a_ 1120 VANAFSFVSQFPQGLDTLVGERGAQLSGGQKQ
                     +ANA  F++ FP G  T+VGERG QLSGGQKQ
                     LANAHGFITSFPNGYQTVVGERGVQLSGGQKQ
H_S09_0013014__-3823 cgagcgtaaatcagtcaggggcggcttggcac
                     tcacagttcgtcagaacttgaggtatcggaaa
                     tacactccttcttataactaaaagagctcggg


CAEEL-PGP-5_1a_ 1152                      RIAIARAILRNPKVLLLDEATSALDS
                                          RIAIARA+L++P +L+LDEATSALD
                                          RIAIARALLKDPPILVLDEATSALDV
H_S09_0013014__-3727                      aagagagccagccacgcgggaagcgg
                     GTATTGG  Intron 4   CAGgtctcgcttaacctttaaccgctat
                     <0-----[3727 : 3670]-0>gttctatcgattatggctagtttatg


CAEEL-PGP-5_1a_ 1178 DSEKV                          VQNALDTASERLSTVVVAHRL
                     +SE++                          V++ALD A +  + +V+AHRL
                     ESERL                          VRDALDRAMKGRTVLVIAHRL
H_S09_0013014__-3591 gagct                          gcggtgcgaagcagtgagcac
                     agagtGTGAGCA  Intron 5   TAGtgactagctaggcttttcagt
                     gtatg<0-----[3576 : 3257]-0>gttcatctgatagcggtccaa


CAEEL-PGP-5_1a_ 1204 STVVNADSIAVLKNGK                              VAEQGTHEEL
                     ST+ NAD I V+K+ K                              V EQGTH++L
                     STIRNADLICVIKDKK                              VQEQGTHDQL
H_S09_0013014__-3193 aaaaaggcatgaagaa                              gcgcgacgcc
                     gctgacattgttaaaaGTCAGTT  Intron 6   CAGtaaagcaaat
                     tcagtctgtctcatgg<0-----[3145 : 3081]-0>gaggtactaa


CAEEL-PGP-5_1a_ 1230 LRKRSIYWRLVQKQGIQVETLIE                       1252
                     +RKR +Y+ LV+ Q        +
                     VRKRGLYYNLVKSQ-------FQDN
H_S09_0013014__-3050 gaacgcttatgatc       tcga                     3003
                     tgaggtaaattaca       taaa
                     cggcgctctgtaaa       cgtt


CAEEL-PGP-5_1a_      *
                     *
                     *
H_S09_0013014__-3002 t                                              3000
                     a
                     g


H_S09_0013014__-2999       [2999 :    1]                                1
                      13 n [2479 : 2467]

//

Gene 1
EXONS 5438 3003
  Exon 5438 5312 phase 0
  Exon 5231 5152
       5151 5150
       5149 5120 phase 1
  Exon 4972 4880 phase 0
  Exon 3841 3728 phase 0
  Exon 3669 3577 phase 0
  Exon 3256 3146 phase 0
  Exon 3080 3003 phase 0
//

Making a G in phase 1 intron


>H_S09_0013014___HAECO_S09_Supercontig_0013014
_5438-5152bp_AA
_5151-5150bp Frameshift
_5149-3003bp_AA
_CAEEL-PGP-1_OR_6_GROUP_CAEEL-PGP-5_1a_1003-
1071AA_1072AA_Frameshift_1073-1252AA

GEVRFENVSFAYPTRPHHQVFECLNLTIPAGQVVALCGPSGEGKSTITSLLERFYTPLSG
RVLLDNQDL

TLNLEWLRGQVIGLISQEPVLFATSIEENIRYGKPSATDEEVMQAARLAN
AHGFITSFPNGYQTVVGERGVQLSGGQKQRIAIARALLKDPPILVLDEATSALDVESERL
VRDALDRAMKGRTVLVIAHRLSTIRNADLICVIKDKKVQEQGTHDQLVRKRGLYYNLVKS
QPQ
DN
//

>H_S09_0013014___HAECO_S09_Supercontig_0013014
_5438-5152bp
_5151-5150bp_Frameshift
_5149-3003bp

ggagaagttcgtttcgaaaatgtaagctttgcatatccgactaggccacaccaccaggtc
ttcgaatgcttgaaccttacaataccagccggtcaagtcgttgcactctgcggaccaagc
ggggaaggtaaaagtacgataacttcactgcttgaacggttctatacgccactctctgga
cgagtgcttttggacaatcaagatctg

ca

acgctgaatttggagtggttgcgagggcaggtcattggattgatttcgcaagaaccagtt
cttttcgctactagtattgaggagaatattcgttacggtaaacccgtcagctacagatgag
gaggtgatgcaagctgctaggcttgcaaacgcacacggtttcatcactagtttccctaat
ggatatcaaacagtcgttggagaacgaggagtgcaattgtccggtggccagaagcagagg
attgctatcgctagagctctcctgaaagatcctccaattctggtgctcgatgaagcgact
agtgctctagatgtggagagtgaacgtttggtgcgtgatgccttagatcgcgctatgaaa
ggtcgaacggtcttggtgattgcccacagactaagtaccataaggaatgccgattctgatt
tgcgttatcaaagatagaaggtgcaagagcaggtacacacgatcaactagtcaggaag
cgcgggctctattacaatttggttaaatcacaattccag
gataat
//
```

A.47 *HAECO*_S09_SUPERCONTIG_0055664 ≈ *CAEEL*-PGP-4/9

```
genewise
Query protein:       CAEEL-PGP-9____CAEEL-PGP-9 C47A10 1 1-1294AA
Target Sequence      H_S09_0055664___HAECO_S09_Supercontig_0055664

H_S09_0055664__     1          Intron     CAG
                               <0-----[     :  112]-0>


CAEEL-PGP-9____   733 VFSNPDRDQMKKDGHFWALMFLVLAAVQGTSMLFQCSLFGVAAERL
                      +FSNPD + +     F ++ FL+L    G +     SLFG+  E++
                      LFSNPDPNALANGNIFNSICFLLLGIGSGITAFASGSLFGITGEKV
H_S09_0055664__   113 tttacgcagcgagaatatattctcgagtgaagtgtgtttgaaggag
                      ttcacacactcagattactgttttgtgcgtcctccgcttgtcgaat
                      gctcgtgtgacttcatcactcgatctatccaacataaactctcaga


CAEEL-PGP-9____   779 TMRIRSKVYR                                      788
                       MR+R  V++
                      AMRLRMDVFK
H_S09_0055664__   251 gactcaggta                                      280
                      ctgtgtatta
                      aggaagtccg
```

```
H_S09_0055664__   281 GTGAAAA  Intron                                 1288
                        <0-----[281  :     ]-0>
//

Gene 1
EXONS 113 280
  Exon 113 280 phase 0
//

>H_S09_0055664___HAECO_S09_Supercontig_0055664
_113-280bp_AA
_CAEEL-PGP-9_733-788AA

LFSNPDPNALANGNIFNSICFLLLGIGSGITAFASGSLFGITGEKVAMRLRMDVFK
//

>H_S09_0055664___HAECO_S09_Supercontig_0055664

ttgttctctaacccggatccgaatgcgctagccaatggtaacatatttaac
tcaatctgtttcctgttacttggcattggatctggcatcacagcattcgcatctggatca
ttattcggtatcactggcgaaaaggtagcaatgcggttacgaatggatgtcttcaag
//
```

A.48 *HAECO*_S09_SUPERCONTIG_0055790 ≈ *CAEEL*-PGP-5/12/14

```
genewise
Query protein:       CAEEL-PGP-14___CAEEL-PGP-14 F22E10 3 1-1327AA
Target Sequence      H_S09_0055790___HAECO_S09_Supercontig_0055790

H_S09_0055790__-2162                        CAG
                     <0-----[     :  587]-0>
                       10 n [1057 : 1048]


CAEEL-PGP-14___   608 SALNNASKGRTTIMIAHRLSTIREADKIVFFEKGVIVEA         646
                       ALN A+KGRTTI+IAHRLSTIR+  KI  +EKG +VE+
                      VALNEAAKGRTTIVIAHRLSTIRDVKKIYVMEKGKVVES
H_S09_0055790__ -586 ggtagggagaaaagagcctaaacggaaatgagagagggtg       469
                      tctaaccaggcctttcagtgctgataatattaagattac
                      gtgtaggaaaacccaattgtttttagaatcgagaaatga


H_S09_0055790__ -468  GTAAGAG  Intron                                  1
                      <1-----[468  :     ]-1>
//

Gene 1
EXONS 586 469
  Exon 586 469 phase 0
//

>H_S09_0055790___HAECO_S09_Supercontig_0055790
_586-469bp_AA
_CAEEL-PGP-14_608-646AA

VALNEAAKGRTTIVIAHRLSTIRDVKKIYVMEKGKVVES
//

>H_S09_0055790___HAECO_S09_Supercontig_0055790

gtggctttgaatgaagcggcgaaaggaagaacaaccatcgtcatagcacatcgt
ttgagtactattcgtgatgtaaagaaaatatatgtcatggaaaagggaaaagtagttgag
tcag
//
```